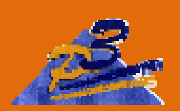
CFPP2000

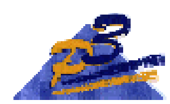


Constitution et exploitation d'un corpus de français parlé parisien

Sonia Branca-Rosoff (U. Sorbonne-Nouvelle, Paris 3-Syled EA 2290)
Serge Fleury (U. Sorbonne-Nouvelle, Paris 3-Syled EA 2290)
Florence Lefeuvre (U. Sorbonne-Nouvelle, Paris 3-UMR Lattice 8094)
Mat Pires (U. Franche-Comté, Laseldi EA 2281)

Mai 2009

SYLED UNIVERSITE PARIS III SORBONNE NOUVELLE



Constitution et exploitation d'un Corpus de Français Parlé Parisien (CFPP2000)

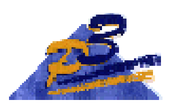
Sonia Branca-Rosoff (U. Sorbonne-Nouvelle, Paris 3-Syled EA 2290)
Serge Fleury (U. Sorbonne-Nouvelle, Paris 3-Syled EA 2290)
Florence Lefeuvre (U. Sorbonne-Nouvelle, Paris 3-UMR Lattice 8094)
Mat Pires (U. Franche-Comté, Laseldi EA 2281)

Résumé

Cet article présente le corpus CFPP2000, composé d'un ensemble d'interviews sur les quartiers de Paris et de la proche banlieue, accessibles sans restriction à l'adresse suivante : http://ed268.univ-paris3.fr/CFPP2000/. Nous abordons successivement les objectifs du projet, les choix opérés pour recueillir et organiser les données, et les outils informatiques qui sont en ligne pour aider les chercheurs à les exploiter. Nous terminons par quelques exemples d'analyse menées à partir de CFPP2000 : le fonctionnement syntaxique de *quoi* périphérique, les commentaires métalinguistiques concernant les parlures populaires, l'usage du couple *par contre/en revanche*, le topos qui rapproche quartier urbain et village.

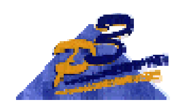
Mots clés

corpus oraux, français parisien, transcription alignée, grammaire de l'oral, espace urbain, langue de la capitale, topos



Sommaire





1. Objectifs de CFPP2000

Convaincue que le développement des technologies de la parole est un enjeu majeur pour les sciences du langage, une petite équipe constituée autour de Sonia Branca-Rosoff, professeure à l'université Paris 3-Sorbonne-Nouvelle (équipe Syled), a entrepris de constituer un corpus échantillonné d'interviews de Parisiens et d'en proposer une transcription synchronisée alignant le texte et le son selon la succession des tours de parole. Outre S. Branca-Rosoff, l'équipe comprend Florence Lefeuvre, maître de conférences (HDR) à l'université de Paris 3-Sorbonne-Nouvelle (équipe Lattice), Mat Pires, maître de conférences à l'université de Franche-Comté (équipe Laseldi), et Serge Fleury, maître de conférences à l'université de Paris 3-Sorbonne-Nouvelle (équipe Syled), qui assure l'organisation du site et le développement d'outils informatiques[1].

Les choix de transcription opérés font de ces corpus des données qui se prêtent bien à des études sur le discours des locuteurs, sur la variation sociolinguistique, sur la syntaxe du français parlé. Pour des recherches faisant intervenir l'intonation, il est toujours possible de revenir au son.

2. Choix opérés dans la constitution du corpus

Il ne peut exister d'échantillon représentatif de « tous les discours en circulation ». Un corpus, s'il est exploitable, est toujours de façon plus modeste le résultat d'un choix correspondant à des échantillons de certains genres discursifs ou de certains locuteurs. Il est donc important d'expliciter nos options, même si nous avons cherché à constituer un corpus qui puisse se prêter à différentes sortes de questionnements.

2.1 L'interview comme genre dialogal et réflexif

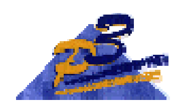
Nous nous en sommes tenus à un genre d'interaction spécifique, l'interview, sur un thème particulier, la relation des personnes interrogées à leur quartier et plus largement à Paris ou à leur banlieue. Nous n'avons donc pas l'ambition de fournir « un corpus de référence » du français parlé en général (Sinclair 1991, Bilger ed. 2000), mais seulement de rassembler des données comparables qui relèvent d'un usage particulier, le français en situation d'interview.

Ce genre nous a paru intéressant parce qu'à la différence du langage lié à l'action, il favorise de longues prises de parole et incite à une attitude réflexive propice au développement d'une syntaxe complexe.

Par ailleurs, le jeu des questions permet d'obtenir une diversité dans les activités linguistiques. Les locuteurs sont ainsi invités à décrire leur quartier, leurs itinéraires, à donner leur opinion argumentée sur plusieurs questions plus ou moins polémiques, à raconter des anecdotes, à s'exprimer de façon métalinguistique sur les langues, particulièrement sur le français, etc.

Nous avons collecté des interviews avec des dyades : la présence de deux personnes face à l'interviewer atténue l'impact d'un observateur extérieur et contribue à

[1] Le Programme de Recherche de la Ville de Paris, dans le cadre de son appel d'offres « Dynamique de l'agglomération parisienne » a fourni les ressources indispensables à l'achat de matériel et au financement des premières transcriptions. Par ailleurs, l'obtention d'une délégation CNRS auprès du laboratoire Praxiling permet à Sonia Branca-Rosoff de se consacrer prioritairement à la réalisation du projet.



détendre l'atmosphère ; elle favorise la discussion, facilite l'expression des désaccords entre les participants ou l'élaboration commune d'opinions. Nous avons aussi recueilli des interviews face à face qui ont suscité des prises de parole plus longues de l'enquêté, l'enquêteur se bornant souvent à l'accompagner par des marques d'approbation (*mh*, *oui*, etc.), sans l'interrompre. Ces interviews offrent un matériau plus facile à exploiter aux spécialistes de prosodie puisqu'elles contiennent peu de chevauchements.

Très souvent les chercheurs cherchent à minimiser leur rôle, à s'effacer complètement. Pour ce projet, qui s'intéresse à la façon dont les personnes collaborent ou s'affrontent pour mettre en mots leurs représentations, cette exigence de neutralité n'a pas paru toujours nécessaire. Nous avons pu chercher à favoriser une discussion peu formelle, ce qui a entraîné par moments une vraie participation des enquêteurs à la conversation.

2.2 Thématique

Nous sommes partis d'un questionnaire portant sur la ville pour plusieurs motifs[2].

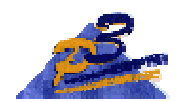
L'intérêt d'un questionnement sur leur pratique de la ville est apparu immédiatement aux enquêtés qui ressentent tous qu'ils ont une expérience à transmettre.

Cette thématique facilite en outre le positionnement des interlocuteurs pendant l'interview. Il est en effet plus facile d'accepter de répondre à des questions sur son cadre de vie que d'accepter un entretien visant à observer la complexité de ses pratiques langagières ou à situer ses productions langagières selon une stratification hiérarchique allant du plus familier (ou du plus populaire), au plus surveillé (ou au plus bourgeois). Il est encore plus difficile, si l'on accepte cette situation inconfortable de s'exprimer de façon libre et spontanée (voir les réflexions de Labov 1976 qui a insisté sur la contradiction entre la situation artificielle créée par la présence d'un observateur et le souci de recueillir le parler vernaculaire des locuteurs).

D'autre part, le thème de la ville a entraîné tout naturellement des questions sur les pratiques plurilingues et sur les traces qu'elles laissent sur le français, qu'il s'agisse des mots provenant de populations historiquement installées depuis longtemps comme les Manouches de Montreuil ou encore des argots professionnels comme l'argot des chiffonniers de Saint-Ouen ou qu'il s'agisse de la présence récente du multilinguisme européen, africain, asiatique. Le questionnaire permet aux citadins de faire le point sur leurs sentiments par rapport à toutes ces langues : essaient-ils de les maintenir et de les transmettre si leurs parents parlent une autre langue que le français, de se les approprier s'il s'agit de leurs voisins ? Des questions sur les parlers des jeunes les amènent à commenter leur stigmatisation éventuelle et à s'exprimer sur le degré de prestige qu'ils attribuent au standard.

Plus généralement le thème choisi correspond à un intérêt pour le rôle éventuel des lieux urbains dans la différenciation linguistique. Cette influence des territoires de la ville peut passer par l'identification à un territoire nettement délimité, que l'on ait affaire à des lieux d'intense brassage de populations, comme c'est le cas pour les banlieues ségrégées de Saint-Ouen ou de Pantin, ou au contraire à des quartiers relativement homogènes comme l'a longtemps été le 7e arrondissement de Paris. Des variétés urbaines peuvent aussi émerger par le biais des réseaux de socialisation (Milroy

[2] Ce questionnaire s'inspire du travail pionnier réalisé à Montréal. (Voir Thibault et Vincent 1990).



et Milroy 1992), paroisses, lycées, groupes de jeunes, habitués de certains cafés, etc. Les phénomènes remarquables peuvent concerner de façon étroite l'adoption de telle ou telle variante, mais aussi l'usage de registres spécifiques plus ou moins consciemment adoptés en fonction de la position du quartier dans l'espace urbain. Une habitante du 11e développe ainsi le thème du rôle structurant de l'opposition est/ouest à Paris. L'opposition qu'elle établit entre les bobos et les bourgeois renvoie selon elle à des thèmes de conversation, autant qu'au « style » relâché qui va avec. Ainsi, l'identification minimale fournie par les paramètres sociologiques, âge, niveau d'étude, profession, etc., est complétée par l'auto-description complexe que les habitants font d'eux-mêmes et des façons dont ils s'inscrivent dans leurs lieux d'habitation.

Au-delà de cet intérêt qui relève de la sociolinguistique classique, nous avons cherché à réfléchir sur la place des quartiers dans la structuration de Paris et de sa banlieue. Le questionnaire comporte des rubriques attendues sur les raisons de la résidence dans le quartier et sur les frontières attribuées à celui-ci.

Il présente des questions sur les réactions des habitants face aux changements intervenus dans la composition de la population, dans l'urbanisme, dans les commerces, etc. Il interroge les habitants des quartiers sur les effets de « la mondialisation » au plan des mélanges de population, des modes culinaires ou vestimentaires. Il leur demande d'évoquer la façon dont ils vivent des situations de « mixité » ou de « ségrégation ».

Ce questionnaire porte encore sur la relation au « temps social » ; le développement de nouvelles formes de fêtes est souvent évoqué dans cette rubrique, fêtes de voisins (d'immeubles), de quartier, nuit blanche, fête de la musique.

L'espace urbain fait également l'objet de questions portant d'une part sur les clivages traditionnels, l'opposition rive gauche, rive droite ; Paris est/ouest, Paris centre/banlieue ; d'autre part, sur les problèmes de déplacements en ville, ce qui conduit les enquêtés à se remémorer l'espace où ils vivent, ainsi que leurs déplacements et à en proposer une expression verbale.

Enfin, l'action politique des municipalités (et de l'Etat) est présente dans beaucoup d'entretiens, souvent introduite par des questions sur la migration, sur l'école, sur la grande pauvreté ou sur la façon de s'informer sur le quartier.

Des historiens, des sociologues, des analystes du discours auront la possibilité d'utiliser ces données dans la perspective de leurs disciplines respectives. Nous évoquons dans la troisième partie de cet article des formes d'exploitation possibles qui se situent dans un champ spécifiquement linguistique.

Par ailleurs, les thématiques abordées sont suffisamment homogènes pour pouvoir mener des comparaisons fines en fonction des zones urbaines, des locuteurs, des relations entre habitants et enquêteurs, etc.

2.3 Zones d'enquêtes

Même si les boulevards périphériques constituent une frontière visible et symboliquement importante, Paris et sa couronne forment à bien des égards une seule agglomération. Nous avons donc inclus quelques communes dans l'échantillon retenu



qui doit pouvoir représenter l'ensemble complexe des variantes du *français de la capitale*[3]. Cependant pour cette première enquête l'est de Paris a été privilégié.

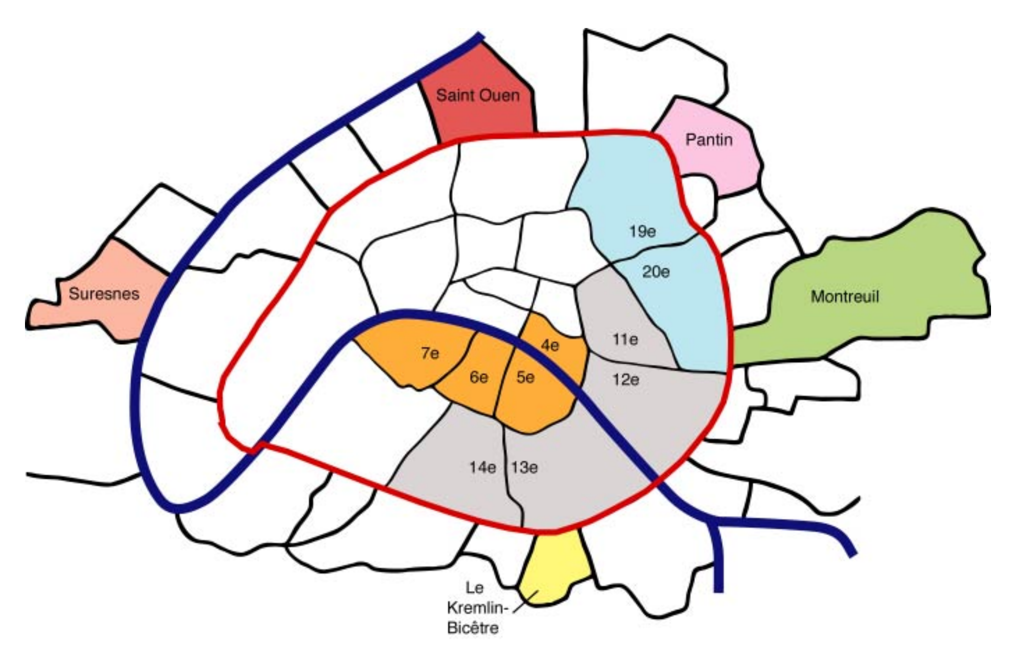


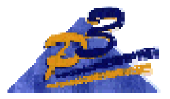
Figure 1 : Les zones d'enquête

Nous avons contrasté quatre zones très approximativement homogènes, quartiers de la grande bourgeoisie, de la moyenne bourgeoisie, jadis populaires et en voie de gentrification, quartiers populaires. Ces zones correspondent à un découpage par communes dans la banlieue et par arrondissements dans Paris. Nous savons que l'unité des arrondissements est en partie fictive, mais les statistiques de l'INSEE[4] sont fournies en fonction de ces découpages administratifs et par ailleurs, les habitants leur accordent une certaine consistance. Nous donnons ci-dessous pour chaque zone, le nombre d'habitants, le prix moyen au m^2 des appartements vendus libres au troisième trimestre 2008[5], le revenu net moyen imposable par foyer fiscal pour 2006, ainsi que les catégories socioprofessionnelles de la personne de référence des ménages interrogés lors du recensement de 1999 (CSP). Les CSP ont donné lieu à beaucoup de discussions et on peut s'interroger sur les frontières qui séparent « les cadres et professions intellectuelles supérieures » et les « professions intermédiaires » ainsi que sur l'homogénéité d'une même catégorie : un employé très diplômé peut se sentir proche d'un cadre et avoir peu en commun avec un employé peu diplômé ou dont le statut est peu valorisé dans l'entreprise. Avec leurs imperfections, ces catégories permettent cependant de dessiner une géographie sociale des différents arrondissements. Les données méta-textuelles permettent aussi de regrouper les textes par niveau de scolarisation, ce qui complète la sous-catégorisation et sera très utile pour les études faisant état de problèmes de "grammaire seconde". Enfin, les enquêtés racontent des

[3] Nous désignons par « français de la capitale » le parler d'un sous-ensemble de Parisiens ayant passé leur enfance dans la capitale. Nous précisons les raisons de ce choix en 2.4.

[4] Institut national de la statistique et des études économiques (http://www.insee.fr/insee-statistique-publique et http://www.statistiques-locales.insee.fr/esl/accueil.asp ;

[5] Chambre des Notaires de Paris, valorisation de l'indice des Notaires 3ème trimestre 2008.



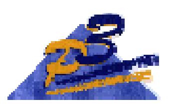
itinéraires de vie qui peuvent remettre profondément en question les socialisations primaires de l'enfance et de l'adolescence, et l'impact de l'univers professionnel.

Les zones d'enquête	**Nombre d'habitants**	**Prix moyen au m² des appartements vendus libres en 2008**	**Revenu net moyen imposable en 2006 (par foyer fiscal)**
4e	30 675 hab.	9 250 €/m2	43 087 €
5e	58 849 hab.	8 650 €/m2	42 079 €
6e	44 919 hab.	10.200 €/m2	73 877€
7e	56 988	9 740 €/m2	90 052 €
11e	149 102 hab.	6210 €/m2	26 122 €
12e	136 591 hab.	6150 €/m2	29 079 €
13e	180 450 hab.	5 950 €/m2	25 525 €
14e	134 700 hab.	6 970 €/m2	32 056 €
19e	187 200 hab.	5 200 €/m2	21 091 €
20e	182 952 hab.	5 400 €/m2	21 220 €
Suresnes (92-Hauts-de-Seine)	39 697 hab.	4 940 € / m²	34 399 €
Kremlin-Bicêtre (94- Val-de-Marne)	23 709 hab.	4100 €/m2	23 070 €
Montreuil (93-Seine-St-Denis)	90 735 hab.	3 730 €/m2	18 422 €
Saint-Ouen (93-Seine-St-Denis)	39 700 hab.	3 230 €/m2	16 007 €
Pantin (93-Seine-St-Denis)	49 919 hab.	2780 €/m2	17 405 €

	Zone grande bourgeoisie				Zone moyen bourgeoisie				Zone en voie de gentrification	
Catégories Socioprofessionnelles	4e	5e	6e	7e	11e	12e	13e	14e	19e	20e
Agriculteurs	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Artisans	5.9	4.6	6.1.	7.0	4.4	3.8	3.3	3.9	4.5	3.7
Cadres, professions intellectuelles supérieures	32.2	32.6	32.2	31.2	24.7	25.7	24.2.	29.3	17.9	18.5
Professions intermédiaires	12.9	12.6	9.3	9.3	17.1	17.6	16.4	15.3	17.0	18.7
Employés	11.6	8.6	8.5	9.3	13.7	13.5	15.3	12.4	17.4	17.3
Ouvriers	3.8	2.8	3.0	3.9	9.1	6.4	7.3	5.2	12.3	11.3
Retraités	21.1	20.5	22.9	24.6	19.7	24.5	23.4	23.3	22.3	22.7
Sans activités	12.6	18.2	18.0	14.7	11.3	8.4	10.2	10.6	8.6	7.8

Figure 2 : Renseignements socio-économique sur les zones d'enquête

Ces résultats sont plus spectaculaires encore si l'on tient compte des évolutions. Selon Michel Pinçon et Monique Pinçon (2004), les cadres moyens, supérieurs et « les patrons de l'industrie et du commerce », sont passés de 35% en 1954 à 65% en 1999 à



Paris, alors qu'inversement, les employés, les ouvriers et les personnels de service qui représentaient 65% des Parisiens en 1954 ne sont plus que 35% en 1999. La pression est particulièrement forte dans l'Est et le Nord de Paris où la population traditionnelle d'ouvriers et d'employés est progressivement remplacée par une population de diplômés du Supérieur. A l'est, les villes de la Petite Couronne qui figurent dans l'échantillon suivent avec retard la même évolution :

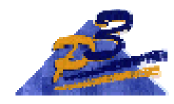
	Zone moyenne bourgeoisie	Zone en voie de gentrification		Zone populaire	
Catégories Socioprofessionnelles	Suresnes	Kremlin-Bicêtre	Montreuil	Saint-Ouen	Pantin
Agriculteurs	0.0	0.0	0.0	0.0	0.0
Artisans	5.0	3.5	3.9	4.7	3.9
Cadres, professions intellectuelles supérieures	24.6	18.3	11.2	7.4	9.0
Professions intermédiaires	16.5	20.5	16.2	13.4	16.1
Employés	16.8	19.0	17.3	20.4	19.9
Ouvriers	9.8	11.5	19.3	22.1	22.8
Retraités	22.8	18.1	25.0	22.5	20.8
Sans activité	4.6	9.1	7.0	9.5	7.5

Pour déterminer les zones, nous avons également tenu compte des représentations traditionnelles sur l'existence d'un parler parisien populaire de Belleville[6] et des communautés linguistiques émergentes influencées par l'émigration à Montreuil, où coexistent des populations issues de l'émigration et des classes moyennes chassées du centre-ville par la hausse des prix de l'immobilier et à Saint-Ouen et Pantin qui amorcent à peine ce processus de gentrification et compensent le déficit de Paris en catégories modestes. Montreuil se caractérise par un mélange ethnique important et par une forte proportion de Maliens, cependant le Bas-Montreuil se peuple depuis les années 80 d'intermittents des arts vivants et de catégories intellectuelles (Collet 2008).

En Seine-Saint-Denis, notamment à Saint-Ouen, la population est une des plus métissées d'Île de France. Espagnols et Italiens arrivés entre 1920 et 1940 coexistent avec les Nord-Africains arrivés après la Seconde Guerre Mondiale, et plus récemment avec les Portugais les Chinois, les Turcs ou les Africains des pays situés au sud du Sahara.

Le choix de zones d'enquête contrastées vise à permettre une étude de la variation linguistique de la capitale. Traditionnellement, on opposait le « français populaire » qui se serait enkysté dans les quartiers modestes ou mal famés de Paris (Gadet 1997) et le « français standard » émergeant d'un lent processus de standardisation du français de la capitale bien décrit par Anthony Lodge (1993, 2004). Le caractère parisien de ce « français standard » apparaît par exemple dans la description donnée par Fouché il y a une quarantaine d'années d'une prononciation française calquée sur « un ensemble de faits phonétiques communs à tous les milieux parisiens cultivés » (1969, p. iii). Cependant ces particularités sont apparues comme très peu marquées dans nos premières interviews. Il nous paraît donc intéressant de montrer à quel point d'unification l'on était parvenu dans la génération des soixantenaires. Une évolution

[6] A cheval sur le 19e, le 20e, le 11e et le 10e.



semble cependant se dessiner par rapport à la notion de français cultivé. On peut la caractériser par une moindre valeur accordée à des variantes puristes. En particulier, au-delà de traits qui relèvent d'un oral ordinaire, le corpus permettra de préciser jusqu'où va le relâchement du respect des injonctions puristes et l'emploi d'un lexique familier, même en situation d'interview. Nous nous intéresserons aussi aux facteurs d'hétérogénéité nouveaux qui se font jour avec la jeune génération des quartiers populaires.

2.4 Catégories de locuteurs

Les locuteurs ont été sélectionnés sur des bases extralinguistiques : un d'entre eux devait être né dans le quartier ou y être venu jeune. Nous ne prétendons pas représenter ainsi la population parisienne. Moins de 25 % des 2 125 246 habitants de Paris recensés en 1999 correspondent à ce critère, d'une part, parce que l'on assiste à un véritable exode des Parisiens, chassés vers la banlieue par le coût des logements qui sont remplacés par des immeubles de bureaux, ou par une bourgeoisie internationale notamment dans les quartiers de l'Ouest de la capitale[7]. D'autre part et simultanément, la part de l'immigration économique s'est accélérée. De 1954 à 1990, la population étrangère est passée de 5 % à environ 15 % de la population parisienne (alors que ce taux est de 5,6 % pour la France entière, et de 11,9 % pour l'Ile-de-France). Aujourd'hui, environ 40 % de la population étrangère vivant en France se trouve en Île-de-France, 57,6 % se concentrant dans les départements de Paris, de la Seine-Saint-Denis et des Hauts-de-Seine. En imposant le critère d'un locuteur au moins ayant passé son enfance à Paris, nous nous donnons ainsi les moyens de situer toutes les variétés qui se croisent dans la capitale par rapport au noyau de base constitué par les Parisiens d'origine. Par ailleurs, les dyades permettent de représenter dans une certaine mesure la réalité sociologique des nouveaux Parisiens venus d'autres régions de France ou de l'étranger.

Dans la suite de nos enquêtes, nous allons essayer de représenter la structure sociale de cette population.

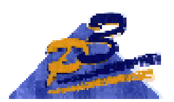
2.5 Taille du corpus

Les corpus transcrits vont d'un peu moins d'une heure à une heure trois-quarts pour les plus longs. En moyenne, ils ont à peu près 20.000 mots.

Il s'agit donc d'enregistrements inhabituellement longs. Cette dimension permet de faire apparaître des variations intéressantes, en fonction des activités langagières pratiquées par les locuteurs et en fonction de l'atmosphère, souvent un peu tendue en début d'interview et beaucoup plus décontractée à la fin.

De longs corpus augmentent aussi les chances de rencontrer plusieurs attestations d'une même tournure, ce qui permet par exemple de caractériser des idiolectes (ainsi les 9 occurrences de *en revanche* pour tout le corpus se rencontrent seulement chez Yvette Audin et une des enquêtrices).

[7] Même si le taux de chômage (au sens du recensement) s'établissait en 1999 à Paris à 12% de la population active au-dessus de la moyenne nationale.



3. Outils

3.1 Outils de génération du corpus

Nous avons voulu avant tout fournir un corpus audio d'une bonne qualité sonore, d'une taille importante, qui soit accessible sur internet et facile à utiliser.

3.1.1 Enregistrement

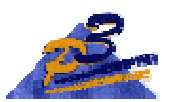
Le matériel utilisé, des enregistreurs *firewire Tascam H2P2*, a permis d'obtenir des enregistrements de bonne qualité. Les fichiers audio ont ensuite été anonymés avec *Audacity*[8]. Tous sont disponibles sous deux formats : WAVE et MP3 (version plus légère, mais de qualité sonore moins bonne).

3.1.2 Transcription

Le logiciel de transcription utilisé est *Transcriber*[9] qui permet de réaliser des transcriptions de l'oral alignées avec le signal. Les raisons de notre choix tiennent à la facilité d'usage de ce logiciel au moment de la transcription et au moment de son utilisation qui doit être possible pour des utilisateurs novices. Les fichiers (nom de fichier ayant pour extension `.trs`) apparaissent sur le site dans le format original produit par *Transcriber* permettant de coder la synchronisation du son et de sa transcription. Le corpus a été découpé au niveau du tour de parole[10].

La **Figure 3** donne à voir l'alignement du fichier audio et de la transcription pour l'une des interviews de CFPP2000 dans l'interface proposée par *Transcriber*. La fenêtre principale contient la transcription découpée en tours de parole. Les boutons sous la transcription permettent de contrôler l'écoute du signal. Celui-ci est visible dans la partie basse de la figure juste au dessus du redécoupage en énoncés (par intervenant). La dernière ligne indique le moment courant (en secondes) depuis le début de l'enregistrement.

8 http://audacity.sourceforge.net/
9 http://trans.sourceforge.net/en/presentation.php
10 Quelques très longs passages ont été découpés en séquences.



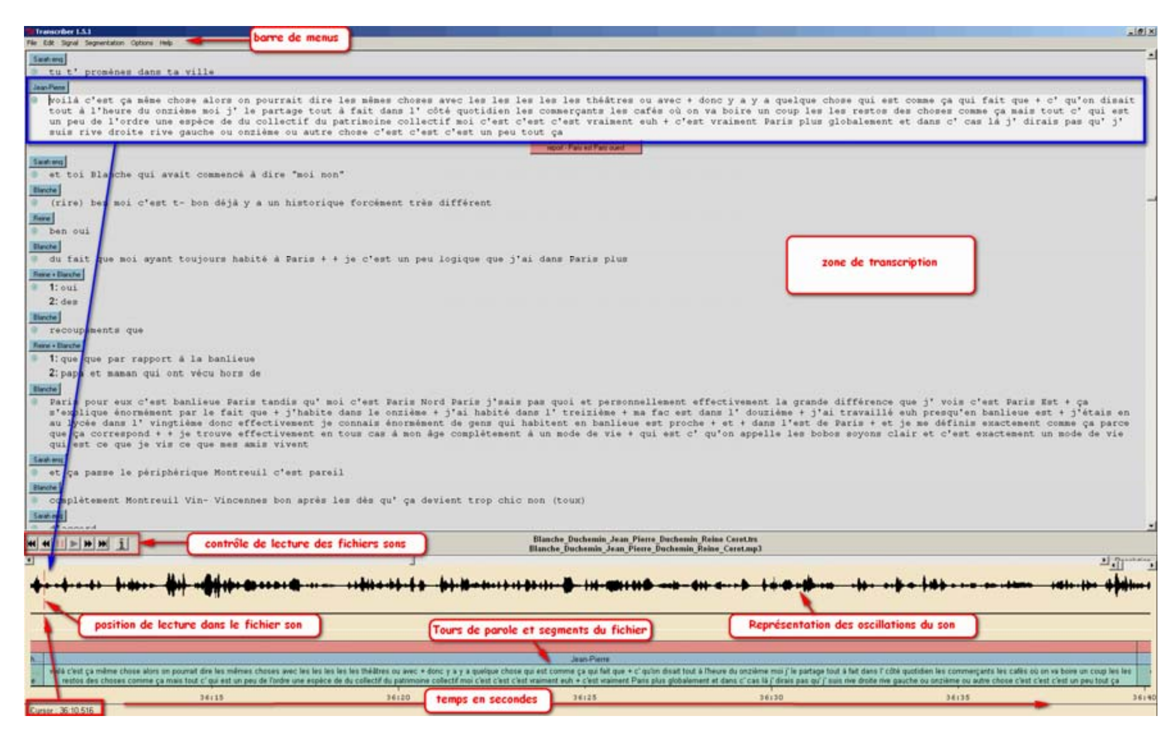


Figure 3 : Interface de Transcriber

La **Figure 4** donne à voir uniquement le fichier de transcription exporté à partir de *Transcriber* (au même « endroit » que dans la figure précédente).




```
<Turn speaker="spk2" startTime="2170.516" endTime="2199.786">
<Sync time="2170.516"/>
voilà c'est ça même chose alors on pourrait dire les mêmes choses avec les les les les les théâtres ou avec + donc y a y a quelque chose qui est comme ça qui fait que + c'
qu'on disait tout à l'heure du onzième moi j' le partage tout à fait dans l' côté quotidien les commerçants les cafés où on va boire un coup les les restos des choses comme
ça mais tout c' qui est un peu de l'ordre une espèce de du collectif du patrimoine collectif moi c'est c'est c'est vraiment euh + c'est vraiment Paris plus globalement et
dans c' cas là j' dirais pas qu' j' suis rive droite rive gauche ou onzième ou autre chose c'est c'est c'est un peu tout ça
</Turn>
</Section>
<Section type="report" topic="to8" startTime="2199.786" endTime="2541.325">
<Turn speaker="spk1" startTime="2199.786" endTime="2202.069">
<Sync time="2199.786"/>
et toi Blanche qui avait commencé à dire "moi non"
</Turn>
<Turn speaker="spk4" startTime="2202.069" endTime="2206.342">
<Sync time="2202.069"/>
(rire) ben moi c'est t- bon déjà y a un historique forcément très différent
</Turn>
<Turn speaker="spk3" startTime="2206.342" endTime="2206.700">
<Sync time="2206.342"/>
ben oui
</Turn>
<Turn speaker="spk4" startTime="2206.700" endTime="2212.950">
<Sync time="2206.700"/>
du fait que moi ayant toujours habité à Paris + + je c'est un peu logique que j'ai dans Paris plus
</Turn>
<Turn speaker="spk3 spk4" startTime="2212.950" endTime="2213.356">
<Sync time="2212.950"/>
<Who nb="1"/>
oui
<Who nb="2"/>
des
</Turn>
<Turn speaker="spk4" startTime="2213.356" endTime="2214.076">
<Sync time="2213.356"/>
recoupements que
</Turn>
<Turn speaker="spk3 spk4" startTime="2214.076" endTime="2215.544">
<Sync time="2214.076"/>
<Who nb="1"/>
que que par rapport à la banlieue
<Who nb="2"/>
papa et maman qui ont vécu hors de
</Turn>
<Turn speaker="spk4" startTime="2215.544" endTime="2255.897">
<Sync time="2215.544"/>
Paris pour eux c'est banlieue Paris tandis qu' moi c'est Paris Nord Paris j'sais pas quoi et personnellement effectivement la grande différence que j' vois c'est Paris Est +
ça s'explique énormément par le fait que + j'habite dans le onzième + j'ai habité dans l' treizième + ma fac est dans l' douzième + j'ai travaillé euh presqu'en banlieue est
+ j'étais en au lycée dans l' vingtième donc effectivement je connais énormément de gens qui habitent en banlieue est proche + et + dans l'est de Paris + et je me définis
exactement comme ça parce que ça correspond + + je trouve effectivement en tous cas à mon âge complètement à un mode de vie + qui est c' qu'on appelle les bobos soyons clair
et c'est exactement un mode de vie qui est ce que je vis ce que mes amis vivent
</Turn>
```


Figure 4 : Extrait de transcription brute



Ce fichier (au format XML) organise dans le temps les tours de parole (marqué ci-dessus par un empilement) par un jeu de jalons textuels (des balises XML) structurant ainsi l'ensemble des informations, il intègre aussi des informations sur les intervenants, les thématiques du discours… La **Figure 5** décrit de manière schématique l'organisation structurelle de ce fichier :

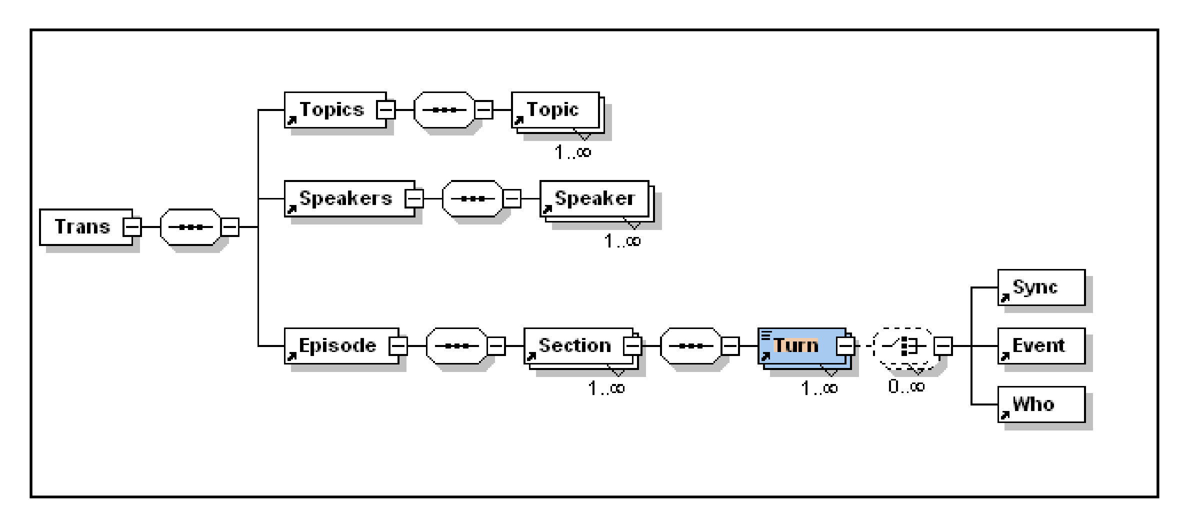


Figure 5 : Arborescence d'un fichier de transcription

Ce schéma décrit de manière plus générale la structure des informations dans un fichier de transcription pour *Transcriber*. Le nœud `Trans` est la racine du document ; on y trouve ensuite des informations générales sur les thèmes, sur les intervenants (nœuds `Topics` et `Speakers`). *Transcriber* permet de segmenter le signal en sections, en tours de parole (dans les éléments associés aux nœuds de la branche inférieure de l'arbre ci-dessus).

Par ailleurs, on peut, dans l'interface de *Transcriber*, faire des recherches de chaîne de caractères dans une transcription, ce qui permet de retrouver facilement n'importe quel passage.

Un mode d'emploi[11] de *Transcriber* est disponible sur le site CFPP2000.

3.1.3 Conventions de transcription

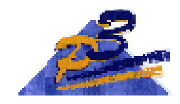
Toute transcription est un compromis forcément boiteux entre le respect des particularités orales et la lisibilité. Puisque le corpus permet de revenir à la version orale, nous avons adopté une transcription orthographique afin de faciliter la lecture cursive et de simplifier l'exploitation des logiciels de concordance.

Le code de transcription adopté est pour l'essentiel celui du DELIC[12] qui a été légèrement adapté.

Nous transcrivons les mots en orthographe sans correction des écarts à la norme quand ils correspondent à un morphème attesté en français.

- Ainsi, bien que ce soit une femme qui s'exprime, nous écrivons *mis* et non *mise* lorsqu'à l'oral la locutrice n'a pas réalisé l'accord du participe :

[11] http://ed268.univ-paris3.fr/syled/ressources/Corpus-Parole-Paris-PIII/tools/Utilisation_de_transcriber.ppt
[12] http://sites.univ-provence.fr/delic/corpus/conventions.html



c'est pas pour des raisons euh catholiques qu'ils m'ont **mis** dans une école de bonnes sœurs (Anita_Musso_11[e], speaker1)

En effet, *mis* est un morphème du français. En revanche, nous n'écrivons pas « f'nêt », « f'nêtr », pour *fenêtre*, ce qui rendrait très difficile les recherches assistées par l'ordinateur dans les corpus.

- Nous avons élidé les clitiques quand cette élision s'entendait à l'oral et écrit *j'* devant consonne, *t'* devant voyelle *: j'sais ; t'arrives*. Ces graphies se sont bien répandues, gagnant peu à peu la bande dessinée, le roman, les blogs. Les adopter amène à produire une description plus proche de la réalité de la morphologie. Il est faux par exemple que la forme sujet et la forme objet direct du pronom de deuxième personne soient en distribution complémentaire, puisque *tu* et *te* sont parfois neutralisés sous la forme *t'*. Cette décision ne devrait pas poser de problèmes puisque les morphèmes « t', j', qu', c', s', d', n' » existent et que la liste en est fermée. En revanche nous nous refusons à « i » ou « iz » pour *ils* qui s'écartent des habitudes orthographiques du français.

- Nous avons respecté les apocopes qui constituent selon nous des néologismes (cela devrait conduire à en tenir compte dans les concordanciers des règles d'équivalence du type : appart = appartement).

- Nous employons des majuscules à l'initiale des noms propres.

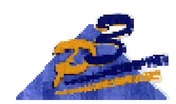
- Le X majuscule note une syllabe incompréhensible.

Amorce, multi-transcription et alternances orthographiques

un mi-	l'amorce d'un mot est notée par un tiret accolé au mot
(d'accord, d'abord)	multi-écoute (hésitation entre plusieurs séquences)
(mais, X)	hésitation entre incompréhension et interprétation
il(s) chante(nt), (ces, ses) on (n') a pas	Les alternances orthographiques n'ont pas été notées systématiquement. Lorsqu'elles le sont c'est entre-parenthèses, conformément à l'usage du DELIC

Prononciation

Dans les premiers corpus, les transcripteurs ont noté (de façon non systématique) par deux-points les allongements remarquables. On a renoncé depuis à ces notations phonétiques.



Liaisons

Les liaisons fautives ont été indiquées par un « z » ou un « t » entre tirets. Ce principe s'étend à des cas où le statut du « z » n'est pas très clair et où il pourrait s'apparenter à un morphème flottant ou à une variante morphologique :

Mille-z-assiettes
donne moi-z-en
des chefs de gare-z-en retraite

Ponctuation

Nous n'utilisons ni les points ni les virgules. Il aurait peut-être été intéressant d'indiquer par une ponctuation forte ou demi-forte toutes les segmentations en unités de discours que le transcripteur perçoit, – que son impression soit due à une pause, par un allongement vocalique ou par un contour intonatif descendant. C. Blanche Benveniste 2007 est d'ailleurs revenue sur la proscription qu'elle avait largement contribué à établir. Nous avons finalement maintenue pour cette première tranche la seule indication des pauses selon les conventions suivantes.

+ ++	pause brève ; longue
///	interruption du discours

- Le point d'interrogation, le point d'exclamation et les guillemets ont été utilisés lorsque le transcripteur entendait nettement l'intonation :

? = interrogations avec montée de la voix
! = exclamation
« » : Les décrochages liés au discours direct sont signalés. Les guillemets ne sont pas toujours fermés, ce qui correspond aux hésitations du transcripteur en ce qui concerne la clôture du discours direct.

Tours de parole

En cas de superposition, les frontières sont posées au niveau des mots de façon à ne pas empêcher leur reconnaissance automatique. De même, lorsqu'un locuteur parle et que le second intervient en arrière-plan en se bornant à des bruits (*mh*) ou à des interjections (*hum*) sans interrompre le tour, ces interventions n'ont pas été notées dans les transcriptions réalisées jusqu'en avril 2009. Elles figurent dans les nouvelles transcriptions à l'intérieur du tour de parole du locuteur principal lorsque celui-ci garde la parole : « ah non moi j'pense que c'est un mixte un peu enfin (*mh*) c'est vraiment c'est pas pauvre c'est pas riche c'est entre les deux ». Les chevauchements sont donc très grossièrement marqués et un retour au son s'impose si l'on veut travailler sur les interactions.



3.1.4 Accessibilité (problèmes juridiques)

Nous avons demandé leur consentement aux locuteurs que nous avons interviewés. Nous leur avons laissé un document mentionnant les finalités de l'enquête et fournissant le nom et l'adresse du responsable de l'enquête.

Les interviewés ont accepté par écrit de nous confier leurs données en sachant que celles-ci devaient être transcrites et mises en ligne. Il a été impossible de leur signaler les types d'utilisation de ces données puisque les corpus sont destinés à être « ouverts » et que dès lors nous ne sommes pas en mesure d'énumérer les types d'exploitation futurs.

Nous avons précisé aux enquêtés que les noms propres des personnes seraient anonymés, sauf en cas de demande expresse de leur part. Nous leur avons également signalé qu'ils pouvaient nous demander de couper certaines parties de l'interview (ce qui a d'ailleurs été parfois le cas).



4. Description de données CFPP2000

Dans cette partie, nous décrivons les ressources accessibles sur la page web du corpus CFPP2000[13]. Les données en ligne regroupent :

(1) les *données initiales* issues des interviews réalisées (fichiers audio) et les fichiers de transcription construits à partir des fichiers audio,
(2) des *fichiers de travail construits* à partir des transcriptions.

4.1 Les données initiales

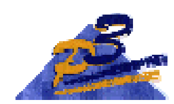
On trouvera en ligne l'ensemble des données issues des interviews réalisées. Pour chacune des interviews, on dispose :

- du fichier audio original (format WAVE),
- d'un fichier audio équivalent mais de taille réduite (format MP3),
- du fichier de transcription (produit avec *Transcriber*),



Figure 6 : Extrait de la page des données en ligne

[13] http://ed268.univ-paris3.fr/CFPP2000/



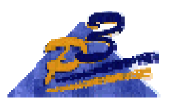
La Figure 6 donne à voir l'allure de la page du site regroupant les données disponibles pour le 11ème. Outre le fichier audio et la transcription de chaque interview, nous avons aussi mis en ligne le fichier de transcription associé à une feuille de styles. Cette version dite « HTML » permet d'afficher « lisiblement » le contenu de la transcription pour quelqu'un qui souhaiterait prendre rapidement connaissance d'un corpus sans charger les fichiers et les programmes (ce qui est nécessaire pour utiliser *Transcriber*). La **Figure 7** donne à voir un extrait du contenu d'un fichier de transcription dans une version paramétrée par une feuille de styles (lien « version HTML » dans la figure précédente).

spk2 [2170.516] : *voilà c'est ça même chose alors on pourrait dire les mêmes choses avec les les les les les théâtres ou avec + donc y a y a quelque chose qui est comme ça qui fait que + c' qu'on disait tout à l'heure du onzième moi j' le partage tout à fait dans l' côté quotidien les commerçants les cafés où on va boire un coup les les restos des choses comme ça mais tout c' qui est un peu de l'ordre une espèce de du collectif du patrimoine collectif moi c'est c'est c'est vraiment euh + c'est vraiment Paris plus globalement et dans c' cas là j' dirais pas qu' j' suis rive droite rive gauche ou onzième ou autre chose c'est c'est c'est un peu tout ça*

spk1 [2199.786] : *et toi Blanche qui avait commencé à dire "moi non"*
spk4 [2202.069] : *(rire) ben moi c'est t- bon déjà y a un historique forcément très différent*
spk3 [2206.342] : *ben oui*
spk4 [2206.700] : *du fait que moi ayant toujours habité à Paris + + je c'est un peu logique que j'ai dans Paris plus*
spk3 spk4 [2212.950] : *oui des*
spk4 [2213.356] : *recoupements que*
spk3 spk4 [2214.076] : *que que par rapport à la banlieue papa et maman qui ont vécu hors de*
spk4 [2215.544] : *Paris pour eux c'est banlieue Paris tandis qu' moi c'est Paris Nord Paris j'sais pas quoi et personnellement effectivement la grande différence que j' vois c'est Paris Est + ça s'explique énormément par le fait que + j'habite dans le onzième + j'ai habité dans l' treizième + ma fac est dans l' douzième + j'ai travaillé euh presqu'en banlieue est + j'étais en au lycée dans l' vingtième donc effectivement je connais énormément de gens qui habitent en banlieue est proche + et + dans l'est de Paris + et je me définis exactement comme ça parce que ça correspond + + je trouve effectivement en tous cas à mon âge complètement à un mode de vie + qui est c' qu'on appelle les bobos soyons clair et c'est exactement un mode de vie qui est ce que je vis ce que mes amis vivent*

Figure 7 : Extrait de transcription

Dans la **Figure 7** les identifiants spk1, spk2… font référence aux interlocuteurs de l'interview. Ceux-ci sont déclarés dans l'en-tête du fichier de description dont le rôle est de donner des informations générales permettant de décrire l'interview. La **Figure 8** donne à voir l'en-tête du fichier de transcription :



Transcription

TOPIC :

1. arrivée dans le 11e
2. quitter la banlieue
3. agréments du onzième
4. mixité du quartier
5. le onzième quartier central
6. les Chinois du onzième
7. limites de la mixité
8. Paris est Paris ouest
9. restaurants cafés
10. les fêtes
11. les commerçants
12. le vélo à Paris
13. transports voitures
14. la propreté
15. les animaux
16. enfance dans le onzième

AUDIO FILENAME :

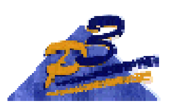
- Blanche_Duchemin_Jean_Pierre_Duchemin_Reine_Ceret.mp3

SPEAKERS :

1. spk1 : Sarah enq
2. spk2 : Jean-Pierre
3. spk3 : Reine
4. spk4 : Blanche

Figure 8 : En-tête de fichier de transcription

Le fichier audio et sa transcription sont accompagnés d'une fiche de description. On donne à voir ici le contenu d'une de ces fiches :



Métadonnées relatives à un enregistrement du Corpus du français parlé parisien des années 2000 (CFPP2000)

Enregistrement
Langue : français
Editeur : Université Paris III – Syled
Date : 29 octobre 2007
Lieu : Université Sorbonne Nouvelle Paris 3, site Censier
Enquêteur : Florence Lefeuvre
Quartier concerné : 4-5-6-7e (7e)
Transcription : Sandy Bichon
Anonymisation : Sonia Branca
Durée : 34'31
Commentaire : Un problème technique a entraîné une brusque interruption de l'enregistrement

Enquêté 1
Etat civil
Prénom et NOM fictif : Laurence Leblond
Sexe : F
Age au moment de l'enregistrement : 43
Situation familiale : mariée, deux enfants
Relation (parenté, ami, etc) avec Enquêté 2 :
Relation (parenté, ami, etc) avec Enquêté 3 :
Réseau par lequel contacté :

Scolarité
Dernier diplôme obtenu : bac+7 : licence d'histoire (3 ans), puis Ecole de commerce extérieur (4 ans)
Commentaires :

Travail
Activité actuelle : responsable marketing et communciation trilingue
Activités passées :

Parents (profession, lieu de naissance, scolarité…)
Mère : secrétaire de banque, née à Nesles-la-vallée (95)
Père : cadre chez Hachette, Antiquariat (livres anciens), né à Agen (47)

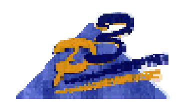
Enquêté 2
Etat civil
Prénom et NOM fictif : Stéphanie Zanotti
Sexe : F
Lieu de naissance, de la petite enfance : 8e ; a habité dans Paris ou région parisienne
Age au moment de l'enregistrement : 49 ans
Situation familiale : mariée, 1 enfant collégienne au collège Jules Romains
Relation (parenté, ami, etc) avec Enquêté 2 :
Relation (parenté, ami, etc) avec Enquêté 3 :
Réseau par lequel contacté :

Scolarité
Dernier diplôme obtenu : bac + 5
Commentaires :

Travail
Activité actuelle : sans profession
Activités passées :

Parents (profession, lieu de naissance, scolarité…)
Mère : née à Besançon
Père : né en Savoie, ingénieur

Figure 9 : Fiche de métadonnées



4.2 Les données construites

4.2.1 Description des corpus lexicométriques

Les données construites sont accessibles sur la page « *Exploration lexicométrique* » du site CFPP2000. Ces données ont été construites pour rendre les contenus textuels de CFPP2000 (le texte des 15 transcriptions) compatibles avec certains outils de traitements automatiques (ici ***Lexico3*** et *Le Trameur* : *cf infra*).

On notera désormais par Corpus-15-cfpp2000[14] le corpus de travail (regroupant les 15 interviews disponibles en mars 2009) constitué à l'issue de la démarche présenté ci-dessous : on retrace ici pas à pas le parcours réalisé (de manière automatique) pour construire ce corpus de travail.

Téléchargement

[V4] : Corpus de Transcription CFPP2000 au format Lexico3/Trameur

Regroupant : Laurence_et_Stephanie_7e, Lucie_da_Silva_F_22_7e, Pauline_de_Bordes_F_67_7e, Raphael_Lariviere_H_23_7e, Yvette_Audin_F_70_7e, Anita_MUSSO_F_46_11e, Mathilde_Lelong_F_85_Marie_Louise_Orsin_F_11e, Blanche_Duchemin_Jean_Pierre_Duchemin_Reine_Ceret_11e, Pierre_Beysson_H_59_Marie_Beysson_F_X_12e, Rosier_Bernard_H_60_Rosier_Micheline_58_12e, Therese_Le_Vern_F_70_Valentine_Testanier_F_60_12e, Nicole_Noroy_F_53_14e, Killian_Belamy_H_22_Lucas_Hermano_H_21_KB, Louise_Liotard_F_85_et_Jeanne_Mallet_75_S0, Andre_Morange_Mo

[V3] : Corpus de Transcription CFPP2000 au format Lexico3/Trameur

[V2] : Corpus de Transcription CFPP2000 au format Lexico3/Trameur

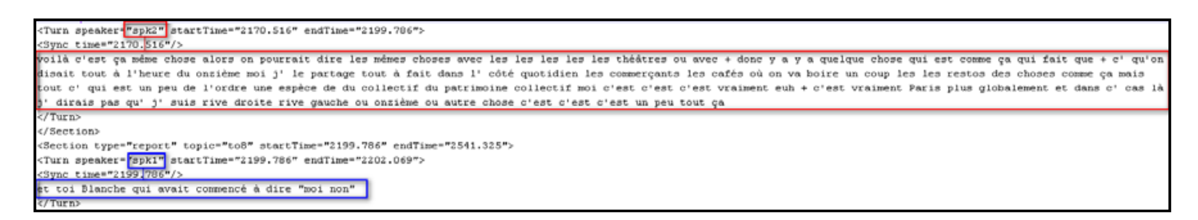
Regroupant : Laurence_et_Stephanie_7e, Lucie_da_Silva_F_22_7e, Pauline_de_Bordes_F_67_7e, Raphael_Lariviere_H_23_7e, Yvette_Audin_F_70_7e, Anita_MUSSO_F_46_11e, Mathilde_Lelong_F_85_Marie_Louise_Orsin_F_11e, Blanche_Duchemin_Jean_Pierre_Duchemin_Reine_Ceret_11e, Pierre_Beysson_H_59_Marie_Beysson_F_X_12e, Rosier_Bernard_H_60_Rosier_Micheline_58_12e, Therese_Le_Vern_F_70_Valentine_Testanier_F_60_12e, Killian_Belamy_H_22_Lucas_Hermano_H_21_KB, Louise_Liotard_F_85_et_Jeanne_Mallet_75_S0, Andre_Morange_Mo

[V1] : Corpus de Transcription CFPP2000 au format Lexico3/Trameur

Regroupant : Laurence_et_Stephanie_7e, Lucie_da_Silva_F_22_7e, Pauline_de_Bordes_F_67_7e, Raphael_Lariviere_H_23_7e, Yvette_Audin_F_70_7e, Anita_MUSSO_F_46_11e, Mathilde_Lelong_F_85_Marie_Louise_Orsin_F_11e, Blanche_Duchemin_Jean_Pierre_Duchemin_Reine_Ceret_11e, Pierre_Beysson_H_59_Marie_Beysson_F_X_12e, Rosier_Bernard_H_60_Rosier_Micheline_58_12e, Therese_Le_Vern_F_70_Valentine_Testanier_F_60_12e, Killian_Belamy_H_22_Lucas_Hermano_H_21_KB, Louise_Liotard_F_85_et_Jeanne_Mallet_75_S0

Figure 10 : Corpus compatibles avec les outils Lexico3, le Trameur et le concordancier en ligne

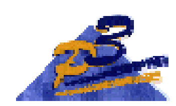
Les données de départ sont les 15 fichiers de transcription construits avec *Transcriber*. Dans ces fichiers, on extrait les données textuelles (les tours de parole) en mémorisant les interlocuteurs et le fichier d'origine (dont le nom est associé à un identifiant d'une transcription). La figure suivante donne à voir un tour de parole dans le fichier de transcription associé à l'interview dont l'identifiant est
`[11-03] Blanche_Duchemin_F_25_Reine_Ceret_F_60_11e` :



```
<Turn speaker="spk2" startTime="2170.516" endTime="2199.786">
<Sync time="2170.516"/>
voilà c'est ça même chose alors on pourrait dire les mêmes choses avec les les les les les théâtres ou avec + donc y a y a quelque chose qui est comme ça qui fait que + c' qu'on
disait tout à l'heure du onzième moi j' le partage tout à fait dans l' côté quotidien les commerçants les cafés où on va boire un coup les les restos des choses comme ça mais
tout c' qui est un peu de l'ordre une espèce de du collectif du patrimoine collectif moi c'est c'est c'est vraiment euh + c'est vraiment Paris plus globalement et dans c' cas là
j' dirais pas qu' j' suis rive droite rive gauche ou onzième ou autre chose c'est c'est c'est un peu tout ça
</Turn>
</Section>
<Section type="report" topic="to8" startTime="2199.786" endTime="2541.325">
<Turn speaker="spk1" startTime="2199.786" endTime="2202.069">
<Sync time="2199.786"/>
et toi Blanche qui avait commencé à dire "moi non"
</Turn>
```

Figure 11: Zoom sur les fichiers de transcription

[14] Disponible en ligne : http://ed268.univ-paris3.fr/syled/ressources/Corpus-Parole-Paris-PIII/work/



Les données textuelles conservées sont : l'interlocuteur (**spk2**) et le contenu textuel, encadré en rouge, du tour de parole, idem pour l'interlocuteur (**spk1**) et le contenu textuel encadré en rouge. L'ensemble des données de ce type pour les 15 transcriptions sont concaténées puis on y insère des jalons textuels (sur la base des informations mémorisées précédemment) pour partitionner l'ensemble et y repérer et caractériser les zones textuelles associées à une transcription donnée.

On rappelle que les identifiants **spk2** et **spk1** font référence aux intervenants de l'interview, on peut retrouver les noms associés à ces identifiants dans l'en-tête du fichier de transcription (*cf* figure *supra*).

Corpus-15-cfpp2000 comportent des jalons textuels permettant de définir plusieurs partitions de l'ensemble du texte *via* les balises `quartier`, `transc`, `user`, `speaker` :

- La balise `quartier` découpe le corpus en parties correspondant aux différents quartiers. Sur l'exemple précédent, l'identifiant est associé au jalon : `<quartier="11">`
- La balise `transc`[15] découpe le corpus en parties correspondant aux différentes transcriptions : `<transc="11-03">` désigne par exemple la troisième transcription de CFPP2000 pour le quartier 11. Sur l'exemple précédent, la transcription dont l'identifiant `[11-03] Blanche_Duchemin_F_25_Reine_Ceret_F_60_11e` correspond à la troisième transcription pour le quartier 11, le jalon inséré est donc : `<transc="11-03">`
- La balise `user` découpe le corpus en parties correspondant aux différents identifiants du corpus. Sur l'exemple précédent, l'identifiant est `[11-03] Blanche_Duchemin_F_25_Reine_Ceret_F_60_11e`, le jalon inséré est donc : `<user="Blanche_Duchemin_F25_Reine_Ceret_F_60_11e">`
- La balise `speaker` découpe le corpus en parties correspondant à chaque tour de paroles. Dans Figure 12, le premier interlocuteur mémorisé (**spk2**) est noté par le jalon : `<speaker="11-03_2">` qui désigne le speaker 2 pour la troisième transcription du quartier 11.

La balise `speaker` est utilisée dans le concordancier en ligne (présenté Figure 15) pour localiser les contextes produits.[16] On présente plus bas une concordance réalisée avec Lexico3 sur la partition `user`.

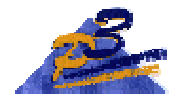
Le caractère (§)[17] délimite ici chaque tour de paroles.

Pour résumer, on présente ci-dessous un extrait de Corpus-15-cfpp2000 donnant à voir les contenus textuels et les jalons définis pour identifier ces parties textuelles :

[15] La balise `transc` est essentiellement une balise « numérique » permettant de distinguer les différents intervenants d'une transcription dans la balise `speaker` décrite *infra*, cette balise étant utilisée dans le concordancier en ligne pour identifier les contextes retournés par cet outil.

[16] Les autres balises peuvent être utilisées comme délimiteur de parties dans les autres outils compatibles avec le format présenté ici.

[17] Ce caractère permet par exemple de visualiser l'ensemble du corpus sous une forme graphique appelée carte de sections, chaque section (un tour de parole) étant matérialisée par un carré. Cette notion est présentée en ligne sur le site de Lexico3 : http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/lexicoWWW/manuelsL3/manuel-3.41.htm#_Toc34045613



Extrait de **Corpus-15-cfpp2000**

```
<quartier="11">
<transc="11-03">
<user="Blanche_Duchemin_F_25_Reine_Ceret_F_60_11e">
…
<speaker="11-03_2">  voilà c'est ça même chose alors on pourrait dire les mêmes choses
avec les les les les les théâtres ou avec + donc y a y a quelque chose qui est comme ça
qui fait que + c' qu'on disait tout à l'heure du onzième moi j' le partage tout à fait
dans l' côté quotidien les commerçants les cafés où on va boire un coup les les restos
des choses comme ça mais tout c' qui est un peu de l'ordre une espèce de du collectif du
patrimoine collectif moi c'est c'est c'est vraiment euh + c'est vraiment Paris plus
globalement et dans c' cas là j' dirais pas qu' j' suis rive droite rive gauche ou
onzième ou autre chose c'est c'est c'est un peu tout ça  §
<speaker="11-03_1">  et toi Blanche qui avait commencé à dire "moi non"  §…
```

Figure 12 : Balises principales utilisées dans le corpus de travail lexicométrique

Le format utilisé ici pour Corpus-15-cfpp2000 correspond à celui utilisé dans les 2 logiciels *Lexico3*[18] et *Le Trameur*[19].

4.2.2 Paramètres lexicométriques de Corpus-15-cfpp2000

Corpus-15-cfpp2000 compte 287 674 occurrences pour 10 646 formes graphiques, 4 617 hapax et la fréquence maximale est de 16 076 (+). La Figure 13 donne à voir les paramètres lexicométriques de la partition *quartier* :

Partie	Occurences	Formes	Hapax	Fmax	Forme
"07"	78119	4999	2486	3837	+
"11"	67841	4561	2230	3800	+
"12"	57877	4111	1990	2627	+
"14"	22281	2068	1081	1068	+
"Kremin-Bicetre"	12755	1601	860	562	+
"Montreuil"	33206	1770	0	3116	+
"Saint-Ouen"	15595	1699	891	1066	+

Figure 13 : paramètres lexicométriques de Corpus-15-cfpp2000

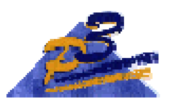
4.3 Les outils CFPP2000

4.3.1 Le concordancier en ligne

Outre *Transcriber*, un prototype de concordancier[20] est disponible sur le site CFPP2000. La figure suivante donne à voir l'interface du concordancier. Pour lancer une requête, il suffit d'entrer la forme graphique cherchée et de définir la longueur du contexte à afficher.

---

18 http://www.cavi.univ-paris3.fr/Ilpga/ilpga/tal/lexicoWWW/
19 http://tal.univ-paris3.fr/trameur/
20 http://ed268.univ-paris3.fr/cgi-bin/search_cfpp2000.cgi



Concordancier CFPP2000

Entrez la longueur du contexte visé (50 par défaut), la forme graphique cherchée (MOT), puis activez la bouton *Recherche*...

50 | MOT | Recherche dans CFPP2000

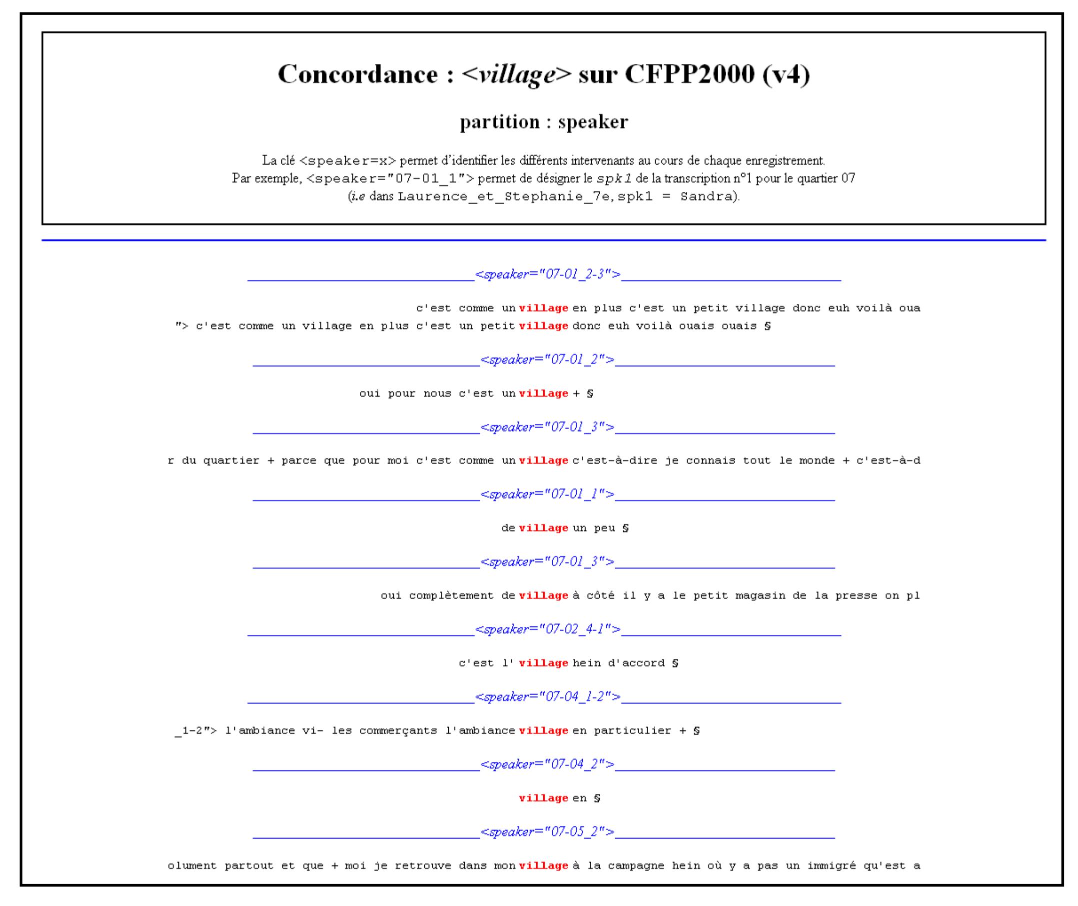
La concordance affichée donnera à voir la partition du corpus permettant d'identifiant les intervenants de la transcription visée.
La clé <speaker=x> permet d'identifier les différents intervenants au cours de chaque enregistrement.
Par exemple, <speaker="07-01_1"> permet de désigner le *spk1* de la transcription n°1 pour le quartier 07
(*i.e* dans [07-01] Laurence_Leblond_F_43_Stephanie_Zanotti_F_49_7e, spk1=Sandra).

La recherche n'est pas sensible à la casse (majuscule/minuscule) : *les* retrouvera *les* ou *Les*.
Il est possible d'utiliser des **expressions régulières**, c'est-à-dire des jokers permettant de remplacer un ou plusieurs caractères :

- le **signe ?** permet de rendre le dernier **caractère facultatif**. Exemple : *les?* cherche *le* ou *les*.
- le **signe .** permet de **remplacer** n'importe quel caractère. Exemple : *m.is* cherche *mais, mois,* etc.
- le **signe** * permet de **répéter** le caractère précédent (qui peut être le joker "." !) de 0 à l'infini. Exemples : *VI** cherche *V, VI, VII, VIII* ; *libéral.** cherche *libéral, libérale, libérales, libéralisation*, etc. (mais pas *libéraux* !).
- le **signe** + permet de **répéter** le caractère précédent de 1 à l'infini. Exemple : *VI* cherche *VI, VII, VIII*, etc. (mais pas *V*)
- le **signe** | permet de spécifier une **alternance**. Exemple: *village|quartier* cherche *village ou quartier*
- les **parenthèses ()** permettent de **grouper** tout ça pour faire des expressions compliquées. Exemples : *(village|quartier)s?* cherche *village, quartier, villages, quartiers* ; *libéra(l|u).** cherche *libéral, libérale, libérales, libéralisation*, etc. et aussi *libéraux*.

Figure 14 : Interface d'interrogation du concordancier

Cet outil permet de construire des concordances sur le fichier Corpus-15-cfpp2000 présenté *supra*. La figure suivante donne à voir un extrait de la concordance de la forme « village » :



Concordance : <*village*> sur CFPP2000 (v4)

partition : speaker

La clé <speaker=x> permet d'identifier les différents intervenants au cours de chaque enregistrement.
Par exemple, <speaker="07-01_1"> permet de désigner le *spk1* de la transcription n°1 pour le quartier 07
(*i.e* dans Laurence_et_Stephanie_7e, spk1 = Sandra).

<speaker="07-01_2-3">

c'est comme un **village** en plus c'est un petit village donc euh voilà oua
"> c'est comme un village en plus c'est un petit **village** donc euh voilà ouais ouais §

<speaker="07-01_2">

oui pour nous c'est un **village** + §

<speaker="07-01_3">

r du quartier + parce que pour moi c'est comme un **village** c'est-à-dire je connais tout le monde + c'est-à-d

<speaker="07-01_1">

de **village** un peu §

<speaker="07-01_3">

oui complètement de **village** à côté il y a le petit magasin de la presse on pl

<speaker="07-02_4-1">

c'est l' **village** hein d'accord §

<speaker="07-04_1-2">

_1-2"> l'ambiance vi- les commerçants l'ambiance **village** en particulier + §

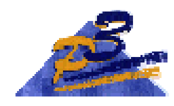
<speaker="07-04_2">

village en §

<speaker="07-05_2">

olument partout et que + moi je retrouve dans mon **village** à la campagne hein où y a pas un immigré qu'est a

Figure 15 : Concordance de *village* en ligne



4.3.2 Les outils d'exploration lexicométrique

Les 2 outils présentés *infra* permettent d'exploiter le fichier Corpus-15-cfpp2000 dans une démarche lexicométrique**.**

Lexico3 [21] est réalisé par l'équipe universitaire **SYLED-CLA2T** (André Salem et alii). ***Lexico3*** est l'édition 2001 du logiciel *Lexico* dont la première version remonte à 1990. Les fonctionnalités présentes dès la première version (segmentation, concordances, décomptes portant sur les formes graphiques, spécificités et analyses factorielles portant sur les formes et les segments répétés) ont été conservées et, la plupart du temps notablement améliorées. L'originalité principale de la série *Lexico* est qu'elle permet à l'utilisateur de garder la maîtrise sur l'ensemble des processus lexicométriques depuis la segmentation initiale jusqu'à l'édition des résultats finaux. Les unités qui seront ensuite automatiquement décomptées sont exclusivement constituées à partir de la liste des délimiteurs fournie par l'utilisateur, sans recours à des ressources dictionnairiques extérieures. Au-delà du repérage des seules formes graphiques, le logiciel permet d'étudier dans les textes la répartition d'unités plus complexes composées de séquences de forme : segments répétés, couples de forme en cooccurrence, etc. au contenu souvent moins ambigu que les formes graphiques dont elles sont composées.

Le ***Trameur***[22] est réalisé par l'équipe universitaire **SYLED-CLA2T** (Serge Fleury et alii). Programme de génération puis de gestion de la *Trame* et du *Cadre* d'un texte (*le métier Textométrique*) pour construire des opérations lexicométriques / textométriques. Le ***Trameur*** intègre le programme *treetagger* : système d'étiquetage automatique des catégories grammaticales des mots avec lemmatisation.

On présente ci-dessous, à titre illustratif, des opérations disponibles dans les 2 logiciels précédents, les résultats ont été construits le fichier **Corpus-15-cfpp2000** présenté *supra.*

4.3.2.1 Concordanciers

Lexico3 et *Le* ***Trameur*** disposent chacun d'un *concordancier*[23]. On donne à voir ci-dessous la concordance de la forme « village » *via* ***Lexico3***; cette concordance permet aussi de visualiser les partitions associées aux contextes affichés (la partition `user` ci-dessous) :

---

[21] http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/lexicoWWW/
[22] http://tal.univ-paris3.fr/trameur/
[23] http://tal.univ-paris3.fr/wakka/wakka.php?wiki=glossairec



```
Partie : "Laurence_et_Stephanie_7e", Nombre de contextes : 6

' est pas impersonnel § c ' est comme un village en plus c ' est un petit village donc euh
omme un village en plus c ' est un petit village donc euh voilà ouais ouais § oui pour nous
à ouais ouais § oui pour nous c ' est un village + § vous vos parents ne n ' habitent pas
er + parce que pour moi c ' est comme un village c ' est - à - dire je connais tout le monde
rès convivial + § ah oui § donc euh § de village un peu § oui complètement de village à côté
 de village un peu § oui complètement de village à côté il y a le petit magasin de la presse

Partie : "Lucie_da_Silva_F_22_7e", Nombre de contextes : 1

e connaît un petit peu mmh § c ' est l ' village hein d ' accord § mmh mmh § pourtant c '

Partie : "Nicole_Noroy_F_53_14e", Nombre de contextes : 3

ous c ' qui fait que c ' était un peu le village quoi parce que on connaissait tout le monde
quand même § ah c ' est me - c ' est mon village hein + ça reste § mon village même maintenant
 ' est mon village hein + ça reste § mon village même maintenant § même si à un moment donné

Partie : "Raphael_Lariviere_H_23_7e", Nombre de contextes : 2

biance vi - les commerçants l ' ambiance village en particulier + § oui § les commerçants
e communauté + + § d ' accord un petit § village en § fait mmh mmh + + § et euh les prises

Partie : "Rosier_Bernard_H_60_Rosier_Micheline_58_12e", Nombre de contextes : 2

 à Saint - Mandé § ouais § ça fait petit village + j ' aime bien § ouais § et : euh § sixième
 § oui § malgré mon : § mon quartier mon village § malgré ma passion pour l ' histoire de

Partie : "Yvette_Audin_F_70_7e", Nombre de contextes : 1

artout et que + moi je retrouve dans mon village à la campagne hein où y a pas un immigré
```

Figure 16 : Concordance de « village » dans ***Lexico3***

D'autres résultats similaires sont disponibles en ligne[24] (concordance, ventilation, mots spécifiques…)

4.3.2.2 Cooccurrents et poly-cooccurrents

*Le **Trameur*** permet de construire des réseaux de cooccurrents de forme pôle (Martinez 2002, 2003, 2006).

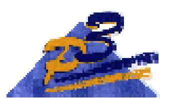
Une *cooccurrence* désigne l'apparition de deux mots en même temps et dans le même contexte.

Le terme *poly-cooccurrence* désigne les attractions lexicales au-delà de la cooccurrence binaire.

La **Figure 17** donne à voir la liste des cooccurrents et le réseau de cooccurrence[25] de la forme « village » sur la partie du corpus réduite au 7e :

[24] http://ed268.univ-paris3.fr/syled/ressources/Corpus-Parole-Paris-PIII/work/

[25] Paramètres du calcul : co-fréquence 2, Seuil : 2, Contexte § (tour de parole). Partie sélectionnée : 7°



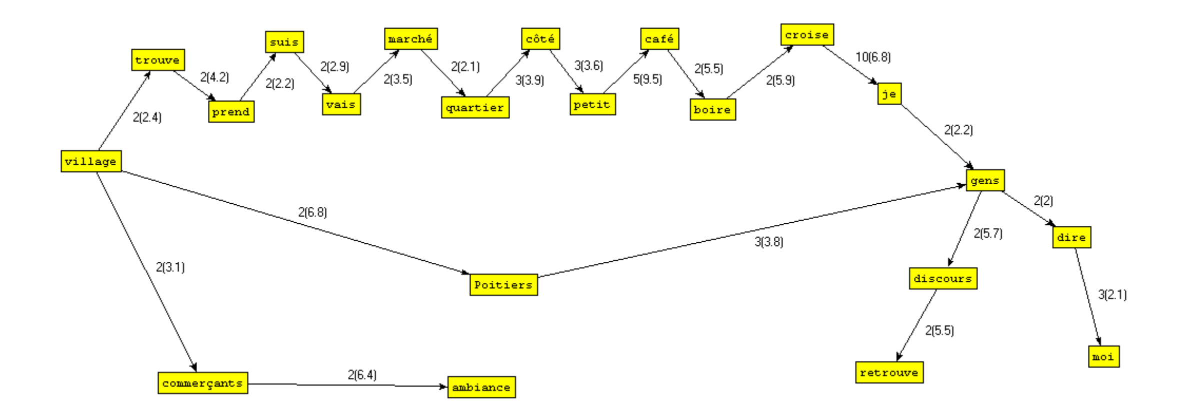


Figure 17 : Poly-cooccurrents de « village » sur la partie du corpus réduite au 7e

4.3.2.3 Etiquetage

Le Trameur permet de construire une base lexicométrique[26] et intègre aussi le programme *TreeTagger*[27] : système d'étiquetage automatique des catégories grammaticales des mots avec lemmatisation. Cet étiquetage permet ensuite de travailler sur les différents niveaux d'annotation disponibles (forme graphique, lemme, catégorie).

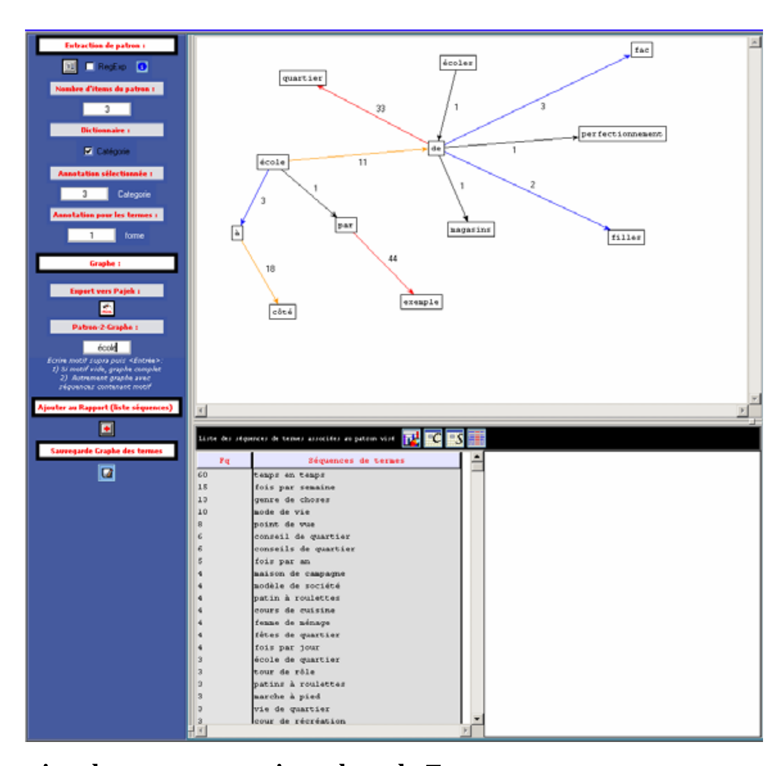


Figure 18 : Extraction de patron syntaxique dans le *Trameur*

[26] *A partir d'un texte segmenté, la numérotation des items découpés dans le texte de départ permet de constituer un système de coordonnées sur le texte dans lequel chaque item est repéré par son numéro d'ordre. Nous appelons ce système de coordonnées sur la séquence textuelle : la **Trame** textométrique.*

[27] http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/DecisionTreeTagger.html



On peut ainsi réaliser une extraction de termes sur la base d'un patron syntaxique donné. La figure qui précède donne à voir :

- la liste des termes construite sur la base du patron `NOM PREP NOM`
- une représentation graphique de l'attraction entre formes graphiques induite par ce patron ; dans l'exemple qui suit le graphe construit est réduit aux termes contenant la forme « école ».

Dans la liste des séquences visibles dans le graphe de la figure ci-dessus on trouve « école de facs » et « écoles de magasins ». Ces 2 séquences « incongrues » correspondent aux contextes suivants (récupérés via le concordancier en ligne) :

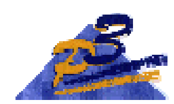
<speaker="07-04_2">

euh y a des petits théâtres euh + euh d'école de **fac** euh §

Dans le second exemple, une frontière syntaxique passe entre *école* et *de magasins de vêtements*.

<speaker="07-02_3">

y avait beaucoup plus euh: à côté d'mon école de **magasins** d'vêtements qui s'ouvraient §



5. Premières exploitations du corpus

5.1 Particularités discursives du français oral : l’exemple de quoi

Dans le prolongement de son ouvrage de 2006, Florence Lefeuvre revient ici sur le fonctionnement de *quoi* périphérique, qui apparaît de façon massive à l’oral[28] :

> *ben ils sont ils sont malpolis* ***quoi*** *euh alors que ils di- ils parleront pas comme ça devant leurs parents (Partie : "12-02_4)*

D’un point de vue syntaxique, nous verrons dans ce *quoi* une forme prédicative dégradée, sorte de “prédicaillon”, renvoyant à ce qui précède, et proche sémantiquement du démonstratif *ça* avec lequel l’affinité de *quoi* est très forte :

> *Malpolis, c’est ça.*

Vu son manque de cohésion syntaxique avec les autres éléments de la phrase, *quoi* a surtout été analysé sur le plan discursif, notamment dans le discours oral où il est le plus récurrent. Fernandez (1994) le range parmi ses particules énonciatives, « articulateurs infimes mais nombreux du discours » (4e de couverture)[29]. Plus précisément, *quoi* est perçu comme jouant un rôle d'« organisation du discours » (Fernandez 1994 : 3). Auchlin (dans Roulet et *al*. 1995 : 93-110) place ce *quoi* parmi les « marqueurs de structuration de la conversation ». D'autres auteurs (Morel et Danon-Boileau 1998) l'analysent comme un ponctuant[30] propre à l’oral.

Notre hypothèse est que *quoi*, de par son sémantisme de non catégorisé, non classifié (cf. Lefeuvre 2005), permet de signaler une recherche dans la formulation des items, de façon massive lorsqu’ils sont de nature rhématique. C’est donc un marqueur intéressant pour repérer les unités syntaxiques rhématiques. Nous reprendrons ici partiellement la définition de Morel et Danon-Boileau (1998) pour qui « le rhème exprime toujours un positionnement singularisé par rapport au jugement que l’on prête à autrui ». Nous ajouterons que ce « positionnement singularisé » marque une nouveauté par rapport aux autres jugements qui ont pu être tenus, conformément à la définition traditionnelle du rhème, opposant le nouveau au connu. Grâce aux outils de lexicométrie, il est possible de procéder à une quantification des données sur une vaste échelle. Cette quantification sera l’objet d’une prochaine recherche.

5.1.1 Contexte linguistique d’emploi

Quoi intervient de façon récurrente dans trois types de contexte.

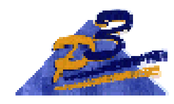
i) Il peut survenir ainsi après un mot dont la formulation est problématique, comme le montre son affinité avec des verbes qui marquent une recherche de l’expression tels que *je veux dire* :

[28] Signalons, sur cette forme linguistique, le travail en cours de Lefeuvre, Morel, Teston-Bernard.

[29] La particule énonciative est définie par plusieurs caractéristiques dans Fernandez 1994 : « brève (généralement monosyllabique), subordonnée prosodiquement à un autre mot ; extérieure au contenu propositionnel de l'énoncé », ce qui les rend facultatives sur le plan syntaxique. Elle est « extrêmement fréquente » à l'oral spontané.

[30] Vincent (1993 : 60) fait remonter l'emploi du terme « ponctuant » à Vendryes (1968 : 174) :

« Le langage parlé ponctue volontiers ses phrases d'une foule de termes dépourvus d'expression, qui fonctionnent comme de la bourre entre les mots expressifs […]. C'est le dernier terme d'une évolution qui les dépouille à la fois de ce qu'elles contenaient d'intellectuel et de sentimental ». (Vendryes, 1968 : 174)



bon + c ' est pas le : de : oui § la rue ***quoi*** *j ' veux dire + c ' est pas (Partie : "12-02_4-5")*
tout l ' monde qui sont des quartiers § sans animation ***quoi*** *j ' veux dire quartiers dits résidentiels (Partie : "07-05_1-2")*

Les répétitions sont alors régulières :

c ' est pas c ' est pas votre votre vie quoi § et j ' m ' en rends compte qu ' on croyait (Partie : "07-03_1-2")

Dans notre corpus qui comporte d'importantes parties descriptives, l'apparition de ce *quoi* est liée aux essais d'explicitation du locuteur :

palier on avait des poêles à des poêles ***quoi*** *pour se chauffer enfin c ' était l ' début (Partie : "Mo-01_2")*

On trouve *quoi* dans des contextes émaillés de marqueurs d'hésitation, comme l'interjection *euh*:

un que j ' connais et puis euh : : voilà ***quoi*** *donc euh § mmh mmh § ça fait bizarre + § (Partie : "07-02_2")*

Ainsi que le signale Vincent (1993 : 98), « l'hésitation fait partie du « parler ordinaire », comme un élément de l'élocution […]. L'hésitation se produit lorsqu'il s'agit de verbaliser la pensée de façon instantanée ; elle fait partie d'un processus en cours. » Vincent (1993 : 97) distingue les ponctuants de régulation, les ponctuants de démarcation, les ponctuants de segmentation et les ponctuants de discours. Dans les ponctuants de régulation, elle y inclut les ponctuants d'hésitation qui « sont émis dans des contextes de trouble ». Elle appelle trouble les « phénomènes liés à la correction ». Le trouble implique une « rupture dans l'énonciation et dans la suite mélodique ». *Quoi* pourrait faire partie de cette catégorie. Mais la réciproque n'est pas vraie. Dès que l'on a un contexte de trouble, *quoi* n'apparaît pas forcément, comme le montre cet exemple tiré de Blanche-Benveniste 1997 :

Ce qui était fantastique dans ce… dans ce… camping, enfin dans ce…, cet hôtel c'était qu'on était carrément en face du Kilimanjaro.

Cela dit, on pourrait l'y insérer :

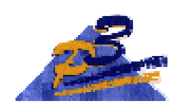
Ce qui était fantastique dans ce… dans ce… camping, enfin dans ce…, cet hôtel ***quoi*** *c'était qu'on était carrément en face du Kilimanjaro.*

ii) *Quoi* peut aussi apparaître dans un contexte d'approximation en s'associant avec des mots tels que l'adverbe *comme* (cf. Le Goffic 1993 : 399) :

c ' est comme une autre façon d ' vivre ***quoi*** *§ mais après on a forcément des + enfin (Partie : "11-03_3")*

Les outils lexicométriques permettent de relever que *voilà* et *quoi* forment une paire lexicale récurrente lorsque *voilà* précède le mot *quoi* :

te dire qu ' les amis euh bon euh voilà ***quoi*** *+ comme on n ' sortait pas § mm § parce (Partie : "12-02_4")*
bon ben parce que on sent que bon voilà ***quoi*** *§ mm § oui c ' est +++ oui oui § XX oui (Partie : "12-02_4")*
mille habitants si vous voulez y a voilà ***quoi*** *mais de de mon point d ' vue ça c ' est (Partie : "Mo-01_2")*
oi j ' crois que j ' suis un peu + voilà ***quoi*** *j ' me suis complètement fermée que ce soit (Partie "11-01_1")*



Voilà, issu de *voir* de par sa nature verbale, joue un rôle prédicatif et permet d’achever (*les amis ? / on sent que ? / y a ?/ j ' suis un peu ?)* le prédicat que le locuteur ne parvient pas à formuler. Son contenu propositionnel se caractérise par un flou sémantique qui peut expliquer la présence de *quoi* après *voilà*. On trouve souvent la configuration *et / puis voilà quoi*, *voilà* formant alors à lui seul un centre prédicatif :

> *un que j ' connais et puis euh : : voilà* ***quoi*** *donc euh § mmh mmh § ça fait bizarre + § (Partie : "07-02_2")*
>
> *ui parce qu ' il est plus large et voilà* ***quoi*** *+ c ' est une question de luminosité aussi (Partie : "12-02_4")*
>
> *e je sens qu ' après ça ramasse et voilà* ***quoi*** *§ mm § ils sont disséminés un peu partout (Partie : "12-02_4")*

Voilà quoi peut s’associer également à *enfin (bon)* qui marque un aboutissement :

> *' habillait comme Johnny on enfin voilà* ***quoi*** *on + vous voyez c ' est + y a un espèce (Partie : "Mo-01_2")*
>
> *llé envoyer en Allemagne enfin bon voilà* ***quoi*** *c ' est un parcours comme beaucoup d ' jeunes (Partie : "Mo-01_2")*

iii) On note enfin la présence régulière avec *quoi* d'un paradigme qui comporte plusieurs segments. Ainsi, dans ces exemples :

> *accord § rue du Commerce jusqu ' au bout* ***quoi*** *§ jusqu ' à une partie donc ça forme un (Partie : "07-02_2")*
>
> *f cent vingt + vingt oui après la guerre* ***quoi*** *+ + § alors on avait donc vous avez § une (Partie : "11-02_2")*
>
> *(1) ux euh Charles Boyer des gens comme § ça* ***quoi*** *mmh mmh § c ' étaient mmh mmh § les (Partie : "11-02_3-2")*

la valeur de variable de *quoi* se déploie par la mise en place du parcours d'une classe d'éléments, dont une partie est explicitée (*rue du Commerce, jusqu’au bout / mille neuf cent vingt, après la guerre / Charles Boyer des gens comme § ça*).

5.1.2 Une recherche de formulation

Ces emplois indiquent que *quoi* signale une recherche de formulation. C’est particulièrement net lorsque se déploie une liste paradigmatique dont les segments correspondent à autant de formulations possibles. Ce type d’emploi pourrait se gloser, pour les exemples ci-dessus, par :

> « Jusqu’au bout, c’est quelque chose comme cela »

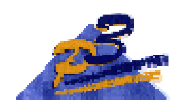
ou alors :

> « Jusqu’au bout, c’est finalement ce que je pourrais dire de mieux de cette limite ; cela s’approche le mieux, mais pas complètement».

Les segments qui appartiennent à un paradigme sont de nature diverse. Ils correspondent à un groupe nominal ou prépositionnel (*rue du Commerce*, *jusqu ' au bout*) ou encore à une proposition relative :

> *§ ah oui oui là là où y a l ' autre voie* ***quoi*** *où y a : où on met les voitures absolument (Partie : "12-01_1")*

Quoi intervient généralement après un mot qui donne une synthèse des formulations antérieures. E. Schepper (1987 : 96) parle d’hyperonyme pour le dernier segment précédant *quoi* (« eine Hyperonymierung ») (*jusqu’au bout / après la guerre / des gens*



comme ça). Dans l'exemple (1), c'est l'interlocuteur qui poursuit la liste paradigmatique en ajoutant le démonstratif *ça*. *Quoi* apparaît avec son pendant, le démonstratif *ça* qui renvoie lui-même à du non encore catégorisé et qui ne dit pas de quoi il s'agit précisément (cf. Corblin 1987).

Plusieurs auteurs signalent cette ouverture de paradigmes. Pour Fernandez, *quoi* fait partie de ces particules qui marquent un piétinement syntaxique (1994 : 177). Celui-ci produit un « inventaire paradigmatique » (*Ibidem* : 178) et s'associe à la « recherche lexicale » (*Ibidem* : 181) avec un souci de précision. Dans Chanet 2001, *quoi* signalerait un ensemble de possibles qui correspondrait à différents paradigmes (paradigme de formulations paraphrastiques, de désignations possibles pour un même référent, de degrés sur une échelle, de voix ou de positions énonciatives, de positions argumentatives).

Quoi signale ainsi un retour du locuteur sur son expression. Par ce processus rétroactif sur ce qui vient d'être dit, *quoi* rappelle les opérateurs de reformulation non paraphrastiques[31], qui n'englobent pas le sème du dire (cf. Rossari 1997). Il permet de « présenter le point de vue introduit comme une reconsidération du point de vue auquel il renvoie » (*Ibidem* : 9). Il introduit un « réexamen » (*Ibidem* : 22) en permettant une prise de « distance modérée » (*Ibidem*) vis-à-vis du point de vue auquel il renvoie. *Quoi* peut être juxtaposé à des reformulatifs non paraphrastiques permettant un réexamen, tels que *somme toute, finalement* :

> *du tout + c ' est pas vraiment euh Paris* ***quoi*** *+ finalement + c ' est + pas y a un esprit (Partie : "07-03_2")*

ou se laisser paraphraser par un de ces reformulatifs :

> *s qui croyaient vraiment en leurs élèves* ***quoi*** *§ mmh mmh + § donc euh + c ' est vrai que (Partie : "07-02_2)*
>
> *s qui croyaient vraiment en leurs élèves* ***finalement*** *§ mmh mmh + § donc euh + c ' est vrai que*

Les reformulatifs non paraphrastiques signalent un « processus de réinterprétation » (Rossari 1997) et *quoi*, en l'occurrence, marque l'appropriation d'un terme et à travers lui d'un point de vue. C'est pourquoi *quoi* peut survenir avec des termes qui marquent le choix d'une opinion tels que le verbe *trouver* :

> *pour ça Paris c'est très pratique en deux minutes on est au cinéma ou dans tel + Galeries Lafayette ou je sais pas quoi voyez enfin moi pour les jeunes je* ***trouve*** *ça génial quoi c ' est sûr que § en fait les jeunes c ' (Partie : "07-01_2")*

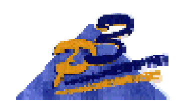
Nous nous rapprochons du constat donné par Fernandez 1994 qui montre que *quoi* est le signe d'une « réappropriation (individuation expressive) de la notion par substitution au terme standard [efficace] d'une formulation familière (ponctuée par la particule *quoi*) » (181-182), comme dans l'exemple suivant :

> *C'est pas… c'est pas vraiment / … comment dire ?… tehokas. C'est pas .. (— efficace)* EFFICACE *voilà. C'est pas cent pour cent quoi* (ex. tiré de Fernandez, 1994 : 181)

J. Fernandez place *quoi* parmi les marqueurs de reformulation paraphrastique se distinguant par leur postposition (*Ibidem* : 176).

A l'inverse de C. Chanet 2001, nous ne verrons pas dans ce mot une invitation du

[31]Cf. pour la distinction entre reformulatifs paraphrastiques et non paraphrastiques, Rossari 1997.



locuteur à l'allocutaire à « partager la subjectivité de son positionnement ». D'après nos données, le rôle par rapport à l'interlocuteur (invitation au partage) distingué par C. Chanet ne semble pas évident, même dans les situations de dialogue. Les deux exemples donnés par l'auteur :

> *L1 nature là ils font trois millilitres ils en font douze donc ils quatre fois plus grands hein en en dimension*
> *L2 comme ceux que l'on achète dans le commerce **quoi non***
> *L1 ben je vais te montrer je dois en avoir un que qui est X là **vois*** (ex. tiré de Chanet 2001)
> *il y a des critères de langage finalement parce que il y a il y a le langage commun quoi et /X, euh/ on s'exprime : comme on s'exprime communément **quoi + tu vois*** (ex. tiré de Chanet 2001)

ne montrent pas que *quoi* est le vecteur d'une connaissance partagée avec l'interlocuteur. En fait, dans ces exemples, ce sont d'autres segments qui font le lien avec l'interlocuteur : *non* et *tu vois*. Nous rejoignons sur ce point M.-A. Morel & L. Danon-Boileau (1998 : 100-102) qui opposent *quoi* et *hein* dans leur valeur de ponctuant[32]. *Hein* s'interprète, selon eux, comme un « appel à l'autre » alors que *quoi* met en avant le centrage sur l'énonciateur parce que, en tant que ponctuant, il est « toujours plat, sans grande intensité ». Il fonctionne comme « un ponctuant fortement égocentré » (*Ibidem* : 103).

Cette recherche de formulation par *quoi* implique-t-elle la validité de l'assertion précédente, comme le signalent M.-A. Morel et L. Danon-Boileau (1998 : 102) ? En fait, d'après le contexte dans lequel survient *quoi*, ce mot n'est pas suffisant pour valider un segment. Deux types de données nous permettent d'avancer cette hypothèse.

Premièrement *quoi* peut être suivi par des segments qui apportent une précision et qui permettent d'affiner la pensée du locuteur. En effet, après l'énonciation de *quoi*, peut survenir un segment qui forme, avec le mot précédant *quoi*, une liste paradigmatique qui ne se trouve pas alors clôturée par *quoi* :

> *où on est dans son réseau euh de classe **quoi** de lycée § alors il s ' trouve que j ' ai (Partie : "07-03_1")*
> *j ' ai déjà vu des chiens enfin costauds **quoi** des pitbulls ou des petits chiens de compagnie (Partie : "KB-01_1-3")*

Cela peut être l'interlocuteur qui poursuit la liste, comme dans cet exemple où est ajouté *jusqu'à une partie* :

> *accord § rue du Commerce jusqu ' au bout **quoi** § jusqu ' à une partie donc ça forme un (Partie : "07-02_2")*

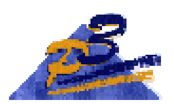
Deuxièmement, on peut remarquer la présence de mots différents de *quoi* qui opèrent une validation de façon explicite. Ces derniers peuvent être exprimés par le locuteur ou l'interlocuteur, à l'aide de *c'est ça* :

> *53.AP. elle a une douleur*
> *54.PA. il est engourdi il est engourdi **quoi***
> *55.AP. il est engourdi voilà c'est ça* (ex. tiré des corpus de Luca Greco)

Dans cet exemple, le locuteur PA répète le segment *il est engourdi* ne trouvant pas de

---

[32]Ils opposent également, avec le même rôle fondamental, le rôle « exclamatif » de *quoi* et de *hein*.



meilleure expression. C'est le locuteur AP qui valide le segment en question. Dans notre corpus, *voilà* permet de valider le mot en question, lorsqu'il suit le mot *quoi*:

> *ces gens qui heu qui ne nous aiment pas* ***quoi*** *voilà + § heu vous n ' trouvez pas qu (Partie : "11-01_1")*
>
> *uoi en gros voilà y avait que le § melon* ***quoi*** *voilà § ou les figues voi - mais le melon (Partie : "12-01_1-2")*

Ainsi *quoi* ne permet pas d'apporter une validation définitive à lui tout seul même si sa position de ponctuant oriente sa portée sur l'élément qui le précède. Il manifeste que l'assertion ne peut pas être de mise telle quelle ; il introduit un doute.
Le contexte linguistique d'emploi montre que *quoi* ne valide pas un segment mais permet de signaler une recherche de formulation.

5.1.3 Marquage d'unités syntaxiques

Cette recherche de formulation ne peut pas concerner n'importe quel mot dans la phrase. Il est difficile pour ce *quoi* de séparer des termes vraiment coalescents, comme une locution verbale du type *avoir besoin* ou bien un clitique et une autre unité, ou encore un déterminant et un nom (cf. Vincent 1993 : 101) :

> **Il a quoi faim*
>
> **Je me quoi trouve dans la salle 408*
>
> **J'ai acheté un quoi journal.*

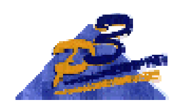
En revanche, comme tout ponctuant[33], il peut démarquer d'autres types de syntagmes, comme un verbe et son COD :

> *et en fait je trouve euh que euh si tu veux d'un côté je préfère* ***quoi*** *euh ne pas avoir d'encadrement tout ça parce que je me sens su-plus libre de mes mouvements quoi* (ex. tiré de Chanet 2001)

C'est pourquoi on peut se demander si *quoi* ne délimiterait pas des unités syntaxiques à l'oral.
Morel et Danon-Boileau (1998) signalent que *quoi* se trouve « à la finale d'un rhème » (p. 102). Les exemples concernant le sujet, a priori thème de la phrase, sont très peu courants (aucun exemple dans notre corpus), même si ce n'est pas impossible. Nous expliquerons l'affinité de *quoi* avec le rhème de la façon suivante : ce qui est l'objet d'une recherche de formulation concerne le plus souvent des segments qui marquent un nouveau jugement dans la phrase, et qui, de ce point de vue, sont rhématiques. Les segments thématiques sont moins concernés par une recherche de formulation. *Quoi* se trouve donc préférentiellement accroché à des segments rhématiques, porteurs de nouvelles informations et, par cela-même, susceptibles d'engendrer une difficulté de formulation. *Quoi* délimite ainsi le rhème de l'énoncé ou du moins une partie rhématique de l'énoncé. Le rhème, dans les énoncés que nous avons pu observer, correspond généralement au prédicat, compris (pour le moment) comme le verbe conjugué principal avec ses compléments, ou à une partie du prédicat de l'énoncé. On peut donc penser que *quoi* permet indirectement de repérer le prédicat de l'énoncé. De fait, dans notre corpus, *quoi* délimite le plus souvent un groupe prédicatif :

---

[33] Pour les critères des ponctuants, cf. par exemple Vincent 1993. *Quoi* n'apparaît presque pas (sauf 1993 : 79) dans l'étude de Vincent 1993 ou Vincent & Dubois 1997 parce qu'il est absent du français québécois.



les autres quartiers on les zappe un peu ***quoi*** *§ et et vous pouviez euh + + essayer d ' (Partie : "Mo-01_2")*

Cela peut être un noyau résumé à *voilà* comme nous l'avons vu précédemment. Cette délimitation indiquée par *quoi* permet d'expliquer également que ce mot puisse se trouver régulièrement à la charnière de tours de parole (signalés par §), c'est-à-dire après la verbalisation du rhème :

blond aux yeux bleus c ' était comme moi ***quoi*** *§ c ' était § Sylvain Weissenberg ++ + y (Partie : "11-03_4")*

is quand on arrivera là ben j ' partirai ***quoi*** *§ quel pessimisme § (rires) voilà (rires) § quel pessimisme § (rires) (Partie : "12-01_1")*

considérables et c ' était am - amusant ***quoi*** *§ mais alors + vous en dites quoi de ces (Partie : "07-03_2")*

Bien sûr la présence de *quoi* étant liée à la recherche de formulation et non au marquage syntaxique proprement dit, le mot *quoi* peut apparaître alors que n'est pas achevée la série paradigmatique :

j ' ai déjà vu des chiens enfin costauds ***quoi*** *des pitbulls ou des petits chiens de compagnie (Partie : "KB-01_1-3")*

ou la partie rhématique :

' trouve que c ' est très inconfortable ***quoi*** *d ' aller fumer sa cigarette dans la rue (Partie : "11-01_1")*

palier on avait des poêles à des poêles ***quoi*** *pour se chauffer enfin c ' était l ' début (Partie : "Mo-01_2")*

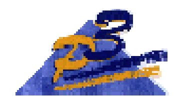
Mais généralement, il fera partie alors, comme ci-dessus, d'un segment du prédicat, celui-ci ayant vocation à être rhématique. *Quoi* peut poser une frontière entre le prédicat et un circonstant extra-prédicatif (*même si c ' est gentiment, parce qu ' on voit les gens qui passent, avec des gens qui viennent de d ' un peu*) :

ins d ' choses à se dire et ça se défait ***quoi*** *même si c ' est gentiment + vous diriez (Partie : "11-01_2")*

i un peu à la fois là un peu à Deauville ***quoi*** *parce qu ' on voit les gens qui passent (Partie : "12-01_1")*

+ j ' me sens enfin mes racines sont là ***quoi*** *+ avec des gens qui viennent de d ' un peu (Partie : "Mo-01_2")*

Ainsi *quoi* signale une recherche de formulation et permet subséquemment de délimiter les unités rhématiques d'un énoncé, qui correspondent le plus souvent aux parties prédicatives des énoncés. Cette hypothèse est à étayer par des données chiffrées, issues de ce corpus.



5.2 Quelques remarques sur la variation

5.2.1 Par contre *et* en revanche : un fait de micro-grammaire

Le discours normatif s'est particulièrement attaché à lutter contre des locutions jugées grossières ou bizarres auxquelles les défenseurs du beau français rêvent de substituer des expressions plus « conformes au « génie de la langue ». Le corpus étant suffisant, Sonia Branca-Rosoff a essayé de mesurer l'impact des prescriptions normatives sur un fait de micro-grammaire caractéristique.

La paire *par contre/en revanche* n'est pas une variable au sens où les sociolinguistes en parlent puisque l'équivalence sémantique des deux lexies est partielle. Elle résulte de l'interdit, auquel est particulièrement associé le nom de Voltaire, à l'encontre de *par contre*, jugement repris par Littré :

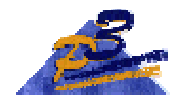
> *Par contre* est une locution dont plusieurs se servent, pour dire en compensation, en revanche : Si les artisans sont ordinairement pauvres, par contre ils se portent bien ; Si le vin est cher cette année, par contre il est bon. Cette locution, qui a été tout particulièrement critiquée par Voltaire et qui paraît provenir du langage commercial, peut se justifier grammaticalement, puisque la langue française admet, en certains cas, de doubles prépositions, *de contre*, *d'après*, etc. mais elle ne se justifie guère logiquement, *par contre* signifiant bien plutôt *contrairement* que *en compensation*, et devant provenir de quelque ellipse commerciale (*par contre* ayant été dit pour *par contre-envoi*) ; en tout cas, il convient de suivre l'avis de Voltaire et de ne transporter cette locution hors du langage commercial dans aucun style. Littré "Contre". Remarque.

On aura noté que le jugement de Voltaire et de Littré, régulièrement repris par la suite, associe la dénonciation d'une expression « bizarre » (la critique des ellipses abusives est très fréquente chez les puristes) et une stigmatisation sociale. Les pratiques langagières des commerçants (comme celles de nombreuses catégories intermédiaires qui incluent particulièrement les fonctionnaires, et aujourd'hui les journalistes…) sont des cibles privilégiées des amateurs de beau langage.

L'Académie, en 1988 a une formule gênée « La locution, décrète-t-elle, ne peut être considérée comme fautive, mais l'usage s'est établi de la déconseiller, chaque fois que l'emploi d'un autre adverbe est possible. ». Même recommandation tarabiscotée chez Péchoin et Dauphin (*Larousse des difficultés*, 1998-2001) qui notent « Naguère critiqué, "par contre" est désormais admis dans le registre courant »… mais persistent « Dans l'expression soignée, en particulier à l'écrit, préférer "en revanche, au contraire, du moins", en fonction du contexte. »

Seulement, les équivalences proposées ne sont pas toujours possibles. *En revanche* a une connotation positive que n'a pas *par contre*. C'est ce qu'avait bien noté André Gide :

> Trouveriez-vous décent qu'une femme vous dise : "Oui, mon frère et mon mari sont revenus saufs de la guerre ; *en revanche*, j'y ai perdu mes deux fils" ? ».



Dans le corpus, on constate l'échec de cette proscription et l'emploi massif de *par contre*.

Seules 9 occurrences de *en revanche* ont été observées. 4 proviennent d'Yvette Audin, une éditrice de 70 ans, élevée dans le 7e arrondissement et 5 d'une des enquêtrices, Sonia Branca-Rosoff, universitaire de 60 ans. Ces deux personnes n'ont pas employé *par contre*. On constate par ailleurs que l'enquêtrice utilise en revanche comme un introducteur d'élément négatif, puisqu'elle l'oppose « à tant mieux », dans un contexte qui évoque la probable disparition des petits commerces dans ce quartier :

> l'impression que les petits commerçants en revanche ou tant mieux peut-être ne tiennent pas (12-01-2)

108 occurrences de *par contre* ont été relevées dans le corpus. Elles sont le fait de locuteurs à haut capital culturel comme de locuteurs qui ont fait peu d'études et ne sont en rien caractéristiques d'un quartier ou d'une tranche d'âge.

Un exemple ne suffit certainement pas pour affirmer que le purisme a reculé. Il invite en tout cas à aller observer systématiquement les zones traditionnellement offertes à la vigilance sourcilleuse des amateurs de beau langage.

5.2.2 L'apport des questions métalinguistiques : Montreuil et ses Roms

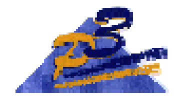
Mat Pires a commencé à dépouiller les questions *sur la langue* qui figurent dans le questionnaire et permettent de compenser un peu l'emploi du standard favorisé par la situation d'enquête.

Les commentaires de la situation linguistique de Montreuil permettent d'illustrer le type de données métalinguistiques issues de ces questions. En effet la question « Dans le quartier, quelles sont les langues qui se parlent ? » suscite dans les deux enregistrements effectués dans cette ville, une évocation d'une même situation linguistique, provoquée par la présence d'une communauté rom.

La communauté rom de Montreuil bénéficie d'une certaine renommée ; son parcours historique et sa présence dans la commune ont déjà fait l'objet d'une monographie (Jaulin, 2000). Les Roms montreuillois sont issus d'une population d'immigrés est-européens établie dans la « zone » des anciennes fortifications à la frontière de Paris, au début de la IIIe République. Suite à la destruction de cette zone par l'Occupant en 1943, une importante population rom se déplace vers Montreuil, ville limitrophe ; d'autres Roms la rejoindront à la Libération de Paris (p. 50-1). Comme le relève le cadre historique affiché du livre de Béatrice Jaulin – il s'intitule *Les Roms de Montreuil, 1945-1975* - la présence des Roms montreuillois, ainsi que leur visibilité sociale, constitue une parenthèse historique qu'accompagne l'embellie économique des Trente Glorieuses ; la remarque de l'écrivain rom Matéo Maximoff, selon laquelle le Montreuil de cette période fut « la capitale des Roms du monde entier »[34] donne la mesure de la présence symbolique de cette communauté. Mais à partir des années 1970 certains membres de la communauté se sédentarisent, se déplacent vers les communes voisines ; d'autres, ne trouvant plus de débouchées dans leurs métiers traditionnels (notamment étamage, chaudronnerie), s'en vont. De nouveaux arrivants, plus récents, ne trouvent plus l'accueil autrefois réservé aux itinérants. Au début des années 2000, des

---

[34] Cité dans Jaulin 2000, p. 16, 53.



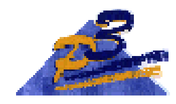
Roms roumains récemment installés sur la commune ont fait l'objet d'un traitement confinant à la persécution de la part des pouvoirs publics (y compris la mairie) : destructions de logement, refus de scolarisation d'enfants, reconduites à la frontière (Lévy-Vroelant et Segal, 2003). Ces évènements, malgré leur aspect négatif, ont toutefois permis aux Roms de rester présents dans les représentations communes de la mixité sociale de Montreuil.

Dans le premier extrait, l'interlocuteur principal est un Montreuillois de 58 ans, André Morange :

Mo-01 André Morange	alors les façons d' parler c'est + à Montreuil oui c'est bon + mais c'étaient euh + des des langages purement montreuillois enfin purement + + y a y a une population d' gitans à Montreuil qui 'est assez connue de Roms + qui avait un langage particulier + qu'on adoptait quand on était jeunes tout simplement +
Sara enq.	il vous en reste des mots ?
André Morange	ouais mais c'est des trucs euh c'est des trucs incompréhensibles maintenant qu' c'est pas
Sara enq.	ah y en a qu' j'connais je sais dire gadjo par exemple
André Morange	oui gadjo mais c'est
Sara enq.	ça c'est assez répandu
André Morange	c'est des choses
André Morange	c'est très répandu ça oui c'est ça + mais + y en a d'autres qui sont + + i- il en connaît une partie Nasser
Nasser	un petit peu un petit peu euh chourave
Sara enq.	chourave oui
André Morange	chourave oui pour voler par exemple
Sara enq.	et on dit pas "j'ai chouravé" on dit "j'ai chourave" + +
André Morange	oh on dit "j'ai chouravé" aussi + +
Sara enq.	ah oui + + d'accord
André Morange	on dit aussi "chouravé" "j'ai chourave j'ai chouravé" euh + + y a plein y a plein d' choses comme ça mais
Sara enq.	ça revient pas parce que
André Morange	qui m'échappent encore qui m'échappent
André Morange	mais c'était vraiment spécifique un peu de + de + de la ville quoi hein de + et ça venait un peu de du enfin de la population manouche quoi de la population

Dans le second, c'est une femme de 67 ans, Marie-Hélène Matera, elle aussi native de la ville :

Mo-02 Marie-Hélène	l'empreinte des: des manouches + parce qu'il y avait beaucoup de manouches et il y en a peut-être encore beaucoup à Montreuil eh + et on a piqué + comme de ci de là quelques mots manouches ou gitans ou je n'sais quoi + par exemple: + qu'est-ce que je pourrais vous citer comme + mais même des fois entre nous on aime bien
Mat (enq.)	ouais ouais ouais bien sûr
Marie-Hélène	employer ce mot par exemple euh on dit: euh euh quelque chose est sympathique vous savez ou quelque chose de bien on dit "mishto" + "mishto" ça veut dire "c'est bien ça" "ça vaut le coup" "c'est + c'est sympa"



Marie-Hélène	on peut pas l' traduire c'est intraduisible
Mat	mm mh + oui oui d'accord
Marie-Hélène	mais « michto » ça reste un mot qu'on emploie
	(…)
Marie-Hélène	un mot + des:: + des "lovés" des "lovés" c'est du c'est des sous + "ah j'ai pas beaucoup de lovés" ça c'est c'est de l'argent vous voyez + euh: + "dicave" ça veut dire "o- regarde" + "o- dicavez dicave euh le gadjo" "dicave le mec dicave le garçon" "le gadjo" c'est un garçon
Mat	oui oui + oui oui
Marie-Hélène	"la gadji" c'est la fille + + voyez + "o- dicave" + ça veut dire ça
Marie-Hélène et Mireille	1. regarde ++ oui on aime bien + parce qu'on aime bien ++ non 2. et vous autres vous preniez un langage vous parliez comme ça parce que vous aviez un petit tsigane avec vous XX ?
Marie-Hélène	non non pas spécialement mais: c'est venu com- comment c'est venu j' peux pas vous dire j' peux pas j' saurais même pas

Les termes proposés, et les remarques métalinguistiques qui les encadrent nous permettent de construire plusieurs analyses, portant sur la perception des langues de l'immigration dans des communautés d'accueil, et sur la valeur sociale des mots empruntés dans la communauté d'accueil, notamment à travers l'évolution sémantique de ces emprunts et de leur réalisation en discours.

Perception des langues de l'immigration

La question posée concerne les *langues* qui se parlent ; les réponses portent donc initialement sur une langue, la langue romanie. Les locuteurs ignorent tout de cette langue. André parle d'un « langage particulier », où l'évitement du mot « langue » a une valeur marginalisante que l'on voit revenir dans sa description des mots romanis comme des « trucs incompréhensibles ». Marie-Hélène, elle, essaye de nommer la langue, mais finit par abandonner : « mots manouches ou gitans ou je n' sais quoi ». Dans ces incertitudes on perçoit le fait que le *romani* – le signifiant comme le signifié - n'est pas reconnu, loin s'en faut, par l'ensemble des locuteurs. Quant au nom du peuple, il est l'objet d'une accumulation de termes dont les nuances échappent généralement aux usagers de la langue (les Roms de Montreuil sont en fait des Kalderash partis de Roumanie au XIXe siècle, cf. Jaulin 2000, p. 34 et suiv.). Cela peut en partie expliquer pourquoi André Morange déplace la description vers la ville d'accueil : parlant d'un « langage purement montreuillois », « vraiment spécifique un peu de + de + de la ville ».

Pour Marie-Hélène Matera la question de la langue est celle de l' « empreinte » des manouches sur sa ville. Cette empreinte semble toutefois relever d'un état passé : là où André parle du phénomène au temps passé (« c'était vraiment spécifique de la ville »), Marie-Hélène doute de l'ampleur actuelle de la communauté : « il y avait beaucoup de Manouches et il y en a peut-être encore beaucoup ».

Mots

Les deux enquêtés proposent quelques exemples de mots romanis. Dans le premier échange, l'apport lexical n'est que de deux mots. Le premier, *gadjo*, est d'abord proposé par l'enquêtrice, tandis que le second, *chourave*, proposé par Nasser, est solidement



lexicalisé en français (*Le Robert* le datant de 1938). Chez Nasser, on relève la disparité entre les connaissances qu'on lui prête - « une partie » des termes – et son apport réel : après avoir minimisé ses connaissances (« un petit peu » au lieu d' « une partie ») il ne propose, finalement, qu'un terme connu. On peut y voir un indice du caractère daté de la présence rom à Montreuil, sans perdre de vue les difficultés que la plupart des locuteurs éprouvent à parler sur les langues.

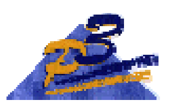
Les termes mentionnés dans le second extrait sont plus nombreux : aux côtés de *gadjo*, on trouve *mishto*, *lovés*, *dicave* - termes absents des dictionnaires courants - complétés de commentaires plus amples, comprenant traductions et exemples d'usage. Toutefois, ces traductions ne sont pas nécessairement neutres. Pour *gadjo*, la définition de Marie-Hélène est des plus banales : *mec* ou *garçon*. L'emprunt a entraîné une neutralisation du sens et des connotations du terme ; on mesure ce glissement à la lecture d'une glose de spécialiste :

> Ce terme est excessivement important et lourd de sens dans la culture rom. Il stigmatise en effet tous les hommes et femmes qui ne sont pas des Roms (...) il est souvent connoté négativement et sous-entend « impur », « dangereux », « différent », « polluant », et « non fiable ».
> L'opposition rom-gadjo met en évidence une frontière clanique, éthique, culturelle, philiosophique et religieuse séparant fortement le peuple des Roms de tous les autres peuples de la Terre (Blanchet 1997, p. 85).

Le mot est donc doté d'une forte valeur symbolique, mais son sémantisme à la fois particulier et vaste (les *gadjé* étant en effet l'ensemble de la population de Montreuil) explique sans doute sa fréquence chez les Roms et sa présence dans le français parlé de cette commune.

Un emprunt d'une langue étrangère subit nécessairement une « naturalisation » lors de son passage d'un code à un autre. La phonologie s'adapte à la langue d'accueil par le remplacement de phonèmes hors-système, par la modification de séquences phonologiques inhabituelles. La morphologie est adaptée notamment dans le cas des désinences (le fait que le nom français n'a pour seule dérivation la marque du pluriel explique en partie la domination nominale de l'emprunt dans cette langue).

Les échanges montreuillois recèlent des exemples de réalisations phonologiques qui peuvent constituer des résistances à la naturalisation des emprunts romanis. Marie-Hélène, en réalisant la séquence qu'on a transcrit en « o- dicave », prononce un *o* ouvert assez curieux, distinct du *o* fermé du mot *oh*. On peut y voir une imitation stéréotypisante de l'exclamation dans la langue de l'emprunt. Il est notable que ce *o* ouvert est même préfixé à la traduction : « o- regarde ». Le terme « dicave » est réalisé de manière systématique avec le phonème /*a*/, ce qu'on pourrait également attribuer à un phénomène d'imitation de la langue d'emprunt. On peut aussi penser à la valeur de ce phonème dans la phonologie du français populaire - il est donné pour supplanté par /a/ en français standard, et surtout dans le français standard de Paris. Le /*a*/ populaire figure dans le répertoire phonémique de la locutrice (cf « des gâteaux des bons gâteaux du bon pain »).



Au niveau de la désinence, l'absence de marque au participe passé dans la séquence *j'ai chourave*, est également notable.[35] Toutefois le fait que cette irrégularité soit mentionnée par l'enquêtrice - l'enquêté, André, tout en acceptant son existence, préférant modérer son propos en rappelant la variante naturalisée *j'ai chouravé* - doit nous conduire à relativiser l'importance de ce type (les données sont bien sûr insuffisantes pour trancher).

Valeur sociale

Dans ces deux cas, il y a un désir apparent de la part du locuteur, d'appuyer l'altérité du vocabulaire utilisé, sa non-appartenance à la langue française. On peut y voir une homologie du regard porté sur les étrangers eux-mêmes, mais on peut aussi interpréter ces préférences dans un cadre *emblématique* (Calvet 1994). Dans ce cas, le choix est porteur d'un sentiment d'identité pour le locuteur, ici d'appartenance à la ville de Montreuil ; l'emploi véhicule une expérience partagé. Les deux enquêtés évoquent explicitement l'aspect identitaire et le caractère plaisant de ces emplois : André parle d'« un langage particulier + qu'on adoptait quand on était jeunes » ; Marie-Hélène indique que « même des fois entre nous on aime bien [employer des emprunts] ». L'aspect emblématique apparaît ici dans le « entre nous » : la locutrice insiste par la suite sur le fait que ces usages survenaient en l'absence de Roms.

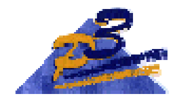
De tels éléments éclairent surtout les rapports entre non-Roms et Roms à Montreuil, de même que dans le livre de Béatrice Jaulin les quelques références à la langue surviennent toutes dans un chapitre consacré à l'intégration locale de la communauté – intégration relativement aboutie étant donné l'attachement à l'entre-soi des Roms. Jaulin évoque ainsi le souvenir de ce « serveur particulièrement zélé qui prenait les commandes en langue tsigane », ou de cette Montreuilloise devenue spécialiste des robes traditionnelles, que la communauté rom tout entière surnommait la Madiska (*couturière* en romani). Les créations de la Madiska auraient de surcroît contribué à pérenniser un métonyme local employé pour les Roms femmes : les « longues robes » (Jaulin 2000, p. 125, 127, 122 et 124). Ce métonyme apparaît en filigrane dans la description des ces femmes que propose Marie-Hélène :

Mo-02 Marie-Hélène	y en a peut-être moins maintenant parce que les femmes on les reconnaît ils sont habillées en longs vous savez ils ont toujours un voile dans les cheveux + un ruban dans les cheveux ça c'est + les robes longues

Autres éléments métalinguistiques

Pour clore cette discussion, rappelons que les questions portant explicitement sur les langues ne représentent pas la seule source d'indications métalinguistiques dans les enregistrements. Ainsi deux indications sur des usages montreuillois (sans rapport avec les Roms) surgissent au cours de la discussion avec Marie-Hélène. Sur le plan lexical, on trouve une référence au « brouillard », un type de plante :

[35] Le seul point de comparaison que l'on peut proposer concerne le verlan. Un verbe comme *pécho* (de *choper*), présente un participe passé invariable : « j'ai pécho ». Toutefois, dans ce cas une désinence invaliderait le statut même de mot verlan, ce qui n'est pas le cas pour *chouraver*.



Mo-02 Marie-Hélène	le quartier
Marie-Hélène	ça a changé à partir du moment + où on a eu les HLM
Mat	ouais
Marie-Hélène	qui ont envahi tous les champs qu'on avait les champs de fleurs + y avait du brouillard où on s' cachait dedans vous savez quand on est gosse le brouillard c'est assez haut + voyez le gypsophile
Mireille	mm hm
Marie-Hélène	nous on appelait ça du brouillard

L'emploi ne figure pas dans les dictionnaires généralistes, mais il est recensé dans les ouvrages spécialisés d'horticulture. Enfin au plan phonologique l'imitation d'une réalisation de d'un /r/ apicale long chez une femme du peuple :

Mo-02 Marie-Hélène	je me rappelle des gens de la rue de: de du passage de XX j'avais des copines bon c'était vraiment des des gens le père il était: éboueur + y avait c'était une famille nombreuse très bien + et la mère elle des fois elle disait faut que j'aille voir le- à la mair::ie qu'elle disait faut que j'aille voir le maire qu'il me donne des bons pour acheter des pâtes pour acheter des trucs comme ça

Au moment d'écrire ces lignes, le corpus ne contient que deux enregistrements effectués à Montreuil ; toutefois les données recueillies et présentées ici permettent déjà de dégager certains aspects de la situation de contact de langues dans cette commune. L'augmentation du nombre d'entretiens devrait nous permettre d'étoffer ces matériaux et ainsi de pouvoir commencer à brosser des portraits linguistiques plus aboutis des quartiers définis.

5.3 Stéréotypie et construction identitaire : « le 7e c'est comme un village »

Dans le prolongement des thèses de Bakhtine sur la circulation des discours (Bakhtine 1984, 1997 et Bakhtine/Voloshninov 1953), Sonia Branca-Rosoff aborde ici des énoncés exprimant des contenus proches et associant le mot *village* à la description d'un quartier. Bakhtine a cherché à comprendre comment chacun vit et s'insère dans un milieu déjà façonné par les discours qui ont précédé sa venue au monde. L'approche bakhtinienne permet d'articuler un travail sur la forme linguistique (ici la répétition du mot *village*), avec la prise en compte de la variation (car bien souvent il s'agit de reconnaître du même, tout en prenant en compte la diversité des formulations.

Ces répétitions intéressent particulièrement l'analyse du discours lorsqu'elles présentent une dimension pragmatique d'acte d'explication ou de justification, et qu'elles montrent comment le discours recoupe l'interdiscours au sens où leur examen permet d'articuler le discours individuel et la dimension collective d'un groupe qui met à la disposition de chaque locuteur des normes et des moyens d'expression. C'est le cas pour ces énoncés qui apparaissent dans des séquences justifiant la préférence accordée au lieu d'habitation.

5.3.1 Les structures attestées

On notera tout d'abord que dans la série recueillie seul un exemple d'emploi du mot *village* réfère au village en tant que petite agglomération campagnarde :



07-05_2 Yvette	dans l'quartier on est affrontés à ce discours là [*le discours raciste*] + il n'est pas + dieu merci universel et y a des tas de gens qui ne l'ont pas [*toux*] mais y a des gens qui l'ont + y a des gens qui l'ont + mais malheureusement c'est un discours qu'on retrouve absolument partout et que + moi je retrouve dans mon village à la campagne hein où y a pas un immigré qu'est arrivé depuis + probablement la bataille de Poitiers euh quand les Arabes sont venus jusqu'à Poitiers

Les autres occurrences du corpus concernent toutes le territoire de la capitale. Elles se répartissent comme suit :

I- Il s'agit surtout de constructions attributives. *Village* attribut peut être :

- précédé d'un déterminant indéfini :

07-01_2 Laurence	en plus c'est un petit village
07-01_3 Stéphanie	oui pour nous c'est un village + §

- précédé d'un déterminant défini :

07-02_2 Francine	tout l'monde se connaît un petit peu mmh c'est l'village hein
07-02_4 Laurence	c'est l'village hein d'accord §

- précédé d'un possessif :

14-01_02 Nicole	ah [*le quatorzième]* c'est me- c'est mon village hein + ça reste mon village même maintenant même si à un moment donné j'aurais dû en- m'en éloigner pour une raison quelconque hein pour euh + euh soit pour le travail ou soit pour le enfin l' travail évidemment mais en tout cas pour le logement

- introduit par d'autres verbes attributifs comme *faire* dans la construction *ça fait* :

12-02_4 Micheline	ça fait petit village + j'aime bien §

II- *Village* apparaît aussi en deuxième position dans des groupes nominaux.

- relié à des substantifs comme *côté* par une préposition :

07-01_2 Laurence	ça a un côté [*enquêtrice* : de village un peu] *Laurence* : oui complètement de village à côté il y a le petit magasin de la presse on plaisante

- employé comme substantif-épithète (Noailly 1990) :

07-04_02 Raphaël	le boulanger + le pois- le poissonnier + enfin vous voyez c'est c'est vraiment l'- l'ambiance vi- les commerçants l'ambiance village en particulier + §

III- Le syntagme comportant *village* peut être accompagné d'adverbes modaux comme *un peu* ou comme *vraiment* dont on sait que, paradoxalement, ils soulignent qu'il ne s'agit pas d'un village au sens propre :



07-04-02 Raphaël	et j'l'ai raccompagnée chez elle et j- j'l'ai un peu rencontrée euh on a parlé + + + voilà donc c'est c'est vraiment une petite communauté + + un petit village en fait
14-01-02 Nicole	euh ils étaient ils avaient la la la poissonnerie rue Daguerre […] euh donc nous c' qui fait que c'était un peu le village quoi parce que on connaissait tout le monde et tout le monde nous connaissait quelque part

- précédé de *comme* employé avec une valeur d'approximation (Fuchs et Le Goffic 2005) :

07-01_2 Laurence	c'est comme un village
07-01_3 Stéphanie	parce que pour moi c'est comme un village c'est-à-dire je connais tout le monde +

IV- Le sujet est un démonstratif (*ce* ou *ça*) qui désigne le référent sans s'embarrasser de propriétés sémantiques. Celles-ci restent un peu floues, même si on trouve souvent le mot *quartier* ou un nom propre dans les tours de parole précédents :

07-01_2 Laurence	non mais en fait ce qui se passe c'est qu'on a bon elles étaient en primaire dans la même école primaire si vous voulez puis après même si on est dans différents collèges **on est toujours dans le même quartier on habite toujours dans le même quartier** donc on se croise forcément on se bon c'est comme ça aussi que nous on est resté amies + on se voit et tout ça + il a des liens qui se créent ce n'est pas impersonnel
07-01_2 Laurence	**c'est comme un village**
07-01_3 Stéphanie	je reviens un petit peu à ce qu'on disait pour la la vie dans le quartier + moi je trouvais enfin bon ça fait juste vingt ans que je suis que j'y suis + **moi je voudrais plus sortir du quartier + parce que pour moi c'est comme un village**
07-02_2 Francine	on a l'impression **dans l'septième on a c'est cocon** parce que à:: si j'vais à Ecole Militaire j'rencontre toujours des amis à moi si j'vais un peu plus bas mmh j'rencontre toujours quelqu'un que j'connais et puis euh:: voilà quoi donc euh mmh mmh ça fait bizarre + tout l'monde se connaît un petit peu mmh **c'est l'village** hein

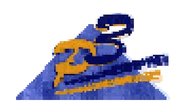
En particulier, le démonstratif renvoie à d'autres catégories que le quartier comme celle du *petit quartier* (explicitement opposé en amont au *quartier-arrondissement*) :

07-04_02 Raphaël	**le petit quartier** alors c'est c'est la rue du Bac + […] ça s'arrête […] à la rue d'l'Université + ça va jusqu'à la rue d'Grenelle […] et puis […] c'est une toute petite fraction mais là pour le coup tout l'monde se connaît + mmh mmh le boulanger+ le pois- le poissonnier + enfin vous voyez c'est **c'est vraiment** […] l'ambiance village

ou comme la catégorie de *l'autre côté* :

xx-xx_2 Rosemonde	c'était pour dire je pense **de l'autre côté** justement c'est plus + peut-être plus vivant voilà **c'est** y a un côté par moment c'est c'est
xx-xx_3 Patrick	**côté village**

Parfois le démonstratif favorise le glissement d'une question faisant référence à un espace de vie à une réponse concernant un mode de vie :



14-01_01 Sonia (enq)	c'était comment euh l'enfance d'un enfant du quatorzième
14-01_02 Nicole	c'était comment euh + ben moi moi j' dois dire bon peut-être que + moi c'était peut-être un petit peu différent d' certains d'autres du fait que mes parents étaient commerçants oui euh donc euh qu'est-ce qu'ils vendaient? euh ils étaient ils avaient la la la poissonnerie rue Daguerre d'accord euh donc nous c' qui fait que c'était un peu le village quoi parce que on connaissait tout le monde

La valeur sémantique de ce démonstratif non classifié permet une référence sans catégorisation (*cf*. Corblin 1987).

Ainsi, *village* n'est pas limité à un type de séquences plus ou moins stable. A côté de nombreux énoncés en *être* où le mot fonctionne comme un attribut qui exprime une qualité ou une propriété que l'on « attribue » au sujet ou qui permet une identification du sujet avec un village, on le trouve aussi en position de complément des noms *côté* (« un côté de village »), *esprit,* ou comme substantif épithète du nom *ambiance*, (« ambiance *village* »), exemple qu'il est possible de paraphraser par *ambiance villageoise*. L'association du quartier et du village à la source des énoncés, ne s'incarne pas dans une expression figée ou en voie de figement. Le rapprochement se situe à un niveau abstrait, mental – cognitif si l'on veut adopter le vocabulaire psychologisant actuel – en tout cas qui opère en amont des énoncés concrets.

5.3.2 Village : une catégorisation à dimension métaphorique

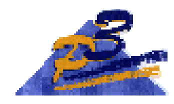
L'interprétation des énoncés rapprochant un lieu d'habitation urbain et un *village* fait cependant intervenir le contexte et ce, à trois niveaux[36]. Le contexte syntaxique met en jeu l'interférence entre le lexique et les constructions grammaticales. Le fonctionnement textuel intervient également puisque l'ambiance villageoise des quartiers est évoquée lors de descriptions argumentatives qui évoquent les avantages de son quartier. Enfin, le contexte historique influe sur les interprétations possibles en éliminant les représentations négatives du village. En effet, alors qu'on a, en principe, la possibilité d'associer aux villages des jugements positifs ou négatifs, l'ambiance-village des quartiers de Paris a toujours été "vue" comme agréable. Pour expliquer ce fait, nous faisons intervenir l'expérience sociale des locuteurs et des auditeurs de ce début du 21e siècle.

- **Le contexte proche : syntaxe et lexique**

D'une part, même si elles peuvent prendre des formes syntaxiques variées, ces séquences poussent toutes à une interprétation métaphorique, qu'il s'agisse des prédications en *être*, des substantifs épithètes, ou des structures N de N.

Village n'a de rôle référentiel dans aucun des énoncés que nous examinons. Il ne comporte pas ses traits habituels de catégorisation sémantique, ceux qui contribuent à permettre l'identification des objets du monde. Par exemple, le trait

[36] Voir J.-M. Barberis 1994 et 1997 pour une vision proche. Nous partageons avec l'approche praxématique que développe cette linguiste l'intérêt pour l'approche linguistique des catégories et le souci d'une contextualisation qui fasse intervenir la dimension sociale des discours.



générique, « agglomération rurale » ne s'applique pas. On ne peut pas non plus le combiner avec des compléments de noms ou des épithètes susceptibles d'y être normalement accolés, pour indiquer par exemple le milieu naturel où se trouve un village particulier ou les activités principales qui permettent de constituer une sous-classe parmi les référents auxquels correspond le nom de village :

Mon village[37] est un village de Provence
? Mon quartier est un village de Paris

Mon village est un village de pêcheurs
? Mon quartier est un village de bureaux

Mon village est un village touristique
? Mon quartier est un village touristique

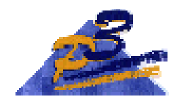
Seules sont au premier plan les propriétés qui conviennent aussi à la description des quartiers urbains. Dans la terminologie de F. Rastier (1987), on peut dire que ce sont les sèmes afférents de *village* qui sont activés. Il ne s'agit pas seulement d'un phénomène associé à la construction attributive. Des effets proches sont obtenus par des groupes nominaux où *village* apparaît en deuxième position comme complément des noms *côté* (« un côté de village »), *esprit* ou comme substantif épithète du nom *ambiance*.

Les substantifs comme *esprit de*, "manière d'agir qui conditionne la vie quotidienne", et *ambiance* "qualité de vie qui conditionne la vie quotidienne" viennent préciser que le substantif village est en quelque sorte vidé d'une partie de ses propriétés et que c'est seulement ce qui touche à la qualité de vie des villages qui est en cause :

07-04_02 Raphaël	pour le coup tout l'monde se connaît + mmh mmh le boulanger: + le pois- le poissonnier: + enfin vous voyez c'est c'est vraiment l'- l'ambiance vi- les commerçants **l'ambiance village**

Par ailleurs, la recherche automatique des cooccurrents de « village » (co-fréquence : 2, Seuil : 2, Contexte : § (tour) sur une partie du corpus exploité (cf **Figure 17**) suggère immédiatement le rôle que jouent dans le caractère convivial d'un quartier les déplacements des habitants (*croiser*, *retrouver*, *vais*, *gens*), les commerces (*marché*, *commerçants*). On note aussi la présence de l'adjectif *petit*. Un retour aux contextes montre le caractère valorisant de *petit* traité comme un hypocoristique qui convient au village dans son entier (« petit village », « petite communauté »), ou qui est appliqué aux commerces de quartier (*petit magasin de la presse*, *petit photographe*). Que ces petits commerces soient associés à l'image du village paraît d'ailleurs piquant lorsqu'on sait que de plus en plus de villageois vont faire leurs courses dans des

[37] Merci à G. Cislaru qui a attiré mon attention sur les problèmes spécifiques posés par les noms propres. Le test permet d'opposer le comportement de « mon village est un village de Provence vs ? Mon quartier est un village de Paris ». Il fonctionne moins bien avec les noms propres : Si « Montmartre est un village de Paris » n'est pas attesté sur le net, on trouve des structures proches comme « Montmartre, le plus célèbre village de Paris se met en fête chaque année grâce aux membres et partenaires de l'association "Un village dans Paris : Montmartre". » (http://www.village-paris- montmartre.com/). De fait, le nom Montmartre est attaché « de façon rigide » à un lieu qui a été un village avant d'être rattaché à Paris et d'en devenir un quartier. Tout se passe comme si le nom propre qui désigne un référent unique conservait toutes les propriétés sémantiques attribuées au lieu (et donc la mémoire de son caractère de village).



hypermarchés. Quoi qu'il en soit, le boulanger, le boucher, le café, le marché sont les médiateurs d'une vie sociale et d'interconnaissances sans implication trop lourde[38].

- **Les types d'activités langagières**

La dimension des types d'activités langagières intervient également dans cet emploi puisque *village* est associé à des activités langagières descriptives et argumentatives. Le motif peut apparaître en fin de séquence pour résumer et renforcer ce qui a déjà été exprimé :

07-01_2 Laurence	puis après même si on est dans différents collèges on est toujours dans le même quartier on habite toujours dans le même quartier donc on se croise forcément on se bon c'est comme ça aussi que nous on est resté amies + on se voit et tout ça + il y a des liens qui se créent **ce n'est pas impersonnel**
07-01_2 Laurence	c'est comme un village en plus c'est un petit village
07-01-3 Stéphanie	donc euh voilà ouais ouais
07-01_2 07-01_3 Laurence	oui pour nous **c'est un village** +

Village apparaît aussi en début de séquence ; l'anomalie qu'il présente – encore soulignée dans l'exemple suivant par le marqueur d'approximation *comme* – est alors levée par le commentaire qui suit immédiatement :

07-01_3 Stéphanie	moi je voudrais plus sortir du quartier + parce que pour moi **c'est comme un village c'est-à-dire** je connais tout le monde + c'est-à-dire que je sors de la maison je vais tous les matins au café + et euh boire le café avec les copines justement les autres mamans des associations des parents d'élèves + euh j'ai un petit photographe à côté qui est adorable on prend le café il y a une galerie à côté d'art on prend le café + euh on croise pas mal de gens euh quand je vais au petit marché aussi je fais le tour du marché on croise souvent des gens « allez viens boire un café » donc je trouve que ça a un côté très convivial +

D'un point de vue textuel, le passage du quartier au village s'apparente à la procédure de reformulation du thème-titre envisagée par Jean-Michel Adam (1992, p. 81) dans le modèle qu'il propose pour la description. Cette « reformulation » est inséparable de l'argumentation, le contexte se chargeant de confirmer l'orientation axiologique nette de village en explicitant les valeurs d'interconnaissance, de convivialité, d'entraide dont le motif est stéréotypiquement porteur[39].

- **Le contexte vécu par les acteurs**

---

[38] Cf. également S. Branca-Rosoff, (2001) qui montre à partir d'une exploration du corpus Frantext qu'une partie des emplois de *quartier* sont associés à ces images de rencontres lors des courses quotidiennes et donc d'interconnaissance « légère » protégeant de la solitude sans créer pour autant d'obligations pénibles.

[39] Surprenants, puisqu'il s'agit d'évoquer la partie du 7e arrondissement située tout près du carrefour très fréquenté Bac/Saint-Germain, sont apparus aussi les motifs du calme et de l'odeur des feux de bois dans les cheminées : « les odeurs de de dès qu'il commence à faire un peu froid de cheminée + euh qu'on allume tout ça on ne le retrouve pas dans tous les + quartiers de Paris » (Patrick Bernard, 7e, dans Rosemonde Gérard et Patrick Bernard).



Enfin la situation dans le monde des interlocuteurs joue un rôle. Les représentations du corpus sont celles d'urbains (par opposition à celles des ruraux), qui vivent au début du 21e siècle.

Ville et *village* constituent depuis longtemps une paire antithétique. Leur opposition est fondatrice des cultures urbaines des derniers siècles. A la ville, synonyme – comme le dit l'étymologie – de *civilité*, s'opposent les villages auxquels ont longtemps été associées des représentations au mieux ambivalentes, ce que montre par exemple une expression proverbiale comme « Cet homme est bien de son village », "Il ne sait rien de ce qui se passe hors d'un cercle très étroit" (*Ac.* 1835) ou « À gens de village, trompettes de bois » pour signifier "Il ne faut aux ignorants, aux grossiers, que des choses proportionnées à leur état, à leur goût, à leur intelligence" (*Ac.* 1798-1878 cité dans Trésor de la langue française). Pour que passent à l'arrière plan les images d'un localisme médiocre, grossier, étouffant et que s'impose massivement l'idée d'un cadre de vie harmonieux et de relations humaines épanouissantes, il a fallu que l'anonymat des grandes villes l'emporte sur les aspects positifs[40] et surtout que les villages se transforment de plus en plus en lieux de villégiature pour des citadins avides de « retour aux sources ». Ainsi le sens n'est pas seulement dans le texte, mais dans les convictions des interprétants du 21e siècle.

5.3.3 Quel statut discursif et sémantique ?

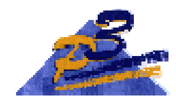
- **Du point de vue de l'analyse des discours**

La notion de *formule* développée par J.-P. Faye dans un livre majeur, réinterprétée dans le champ de la linguistique par M. Ebel et P. Fiala et reprise brillamment par A. Krieg-Planque (2009), a été proposée pour traiter de lexies ou de mots qui s'imposent à un moment donné comme une sorte de passage obligé des discours. Nous n'y aurons pas recours pour plusieurs raisons. Tout d'abord les « formules », (« Etat-total », « mondialisation », « camp d'extermination », etc.) sont des supports de polémiques, alors que les objets dont nous traitons ne donnent pas lieu à débat. Ce sont des mots quasi invisibles, qui ne concentrent pas d'enjeux politiques décisifs et servent à dire le consensus.

Nous ne renverrons pas non plus aux familles d'énoncés étudiées par Michel Foucault dans *Les Mots et les choses* qui relèvent des institutions de savoir et de pouvoir.

La formule correspond à des « moments discursifs » (Moirand 2004) où son emploi s'intensifie brusquement et l'énoncé de Foucault fait événement. Les récurrences dont nous parlons obéissent à une temporalité très différente : elles se propagent de façon plus continue dans les discours quotidiens. Elles relèvent d'un fonctionnement banal que l'on retrouve derrière de très nombreuses unités lexicales qui participent de représentations partagées, peu conflictuelles, même si chacun les actualise à sa façon. Elles appartiennent au langage de tous les jours et entreraient facilement dans le *Dictionnaire des idées reçues* de Flaubert. C'est à l'articulation de la

[40] On ne se hasardera pas à proposer une date et Boileau qui se lamentait sur les embarras de Paris plagiait déjà Juvénal : « en quelque endroit que j'aille, il faut fendre la presse/ D'un peuple d'importuns qui fourmille sans cesse » (Satire VI, 1666).



sémantique lexicale et de l'argumentation qu'il faut chercher les clés de ce fonctionnement.

Les énoncés comportant *village* se rapprochent beaucoup des *topoï* dans la définition qu'en proposent Jean-Claude Anscombre (1995, 2001) et Oswald Ducrot (1995) ou du moins, ils pourraient facilement entrer dans des scénarios argumentatifs du type « plus un lieu se rapproche d'une forme de vie villageoise, plus on s'entraide », etc. La particularité de ces énoncés est cependant qu'ils ne se présentent pas comme des énoncés doxaux. Ils n'en ont pas le caractère sentencieux. Ils sont au contraire présentés comme une vérité constatée en quelque sorte directement par les locuteurs qui peuvent même modaliser leurs assertions pour mieux en montrer l'aspect subjectif (« pour nous, c'est un village »), même s'il serait facile de dégager des enchaînements argumentatifs sous-jacents comme « le quartier est un vrai village, l'atmosphère est donc conviviale », etc.).

- **Du point de vue de la sémantique lexicale**

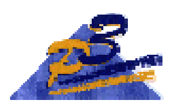
Le couple *quartier/village* est un « organisateur cognitif » qui participe à la construction du sens discursif des énoncés. Symétriquement, les effets du contexte discursif conduisent à revenir sur la signification de l'unité lexicale *village*.

La présentation proposée jusqu'ici a admis la vue référentialiste selon laquelle un nom comme *village* a d'abord pour fonction de désigner un groupe de maisons situées à la campagne. La signification de ce mot serait alors peu à peu déplacée par le contexte dans le sens positif que nous avons évoqué. A terme, *village* pourrait finir par en être affecté durablement jusqu'à condenser la série des formulations positives qui lui sont associées[41] quand il s'agit d'évaluer positivement une zone d'habitation, en particulier urbaine et ce, au prix d'une perte sémantique – la perte de sa fonction première.

En renversant la perspective, d'une façon où pourraient se retrouver des « culturalistes » comme George Lakoff et Mark Johnson (1985), des spécialistes de l'argumentation comme Jean-Claude Anscombre et Oswald Ducrot, des sémanticiens comme Pierre Cadiot et François Nemo (1997), des praxématiciens comme Paul Siblot, (2001), (2007), on pourrait cependant mettre au cœur de l'approche les types de contact qu'entretiennent les hommes avec leur environnement. Tous ces chercheurs ont en effet, quelles que soient les différences importantes qui les séparent, la conviction que notre appréhension du monde est fondamentalement relationnelle : nous percevons les choses du monde à partir des propriétés relatives à notre interaction avec ces choses : « lorsque nous faisons l'expérience du monde, notre culture est déjà présente, dans l'expérience elle-même (Lakoff et Johnson 1985, p. 66).

Nous rejoignons les remarques de D. Legallois 2007, p. 45-46, qui réfléchissant sur la valeur symbolique du village gaulois d'Astérix constate qu'il s'agit d'un objet « appréhendé dans une certaine expérience, cette expérience même que signifie le mot *village*. *Village*, dans une sémantique qui révèle le fondement expérientiel de la signification, manifeste une agglomération d'éléments qui de ce fait constitue un tout homogène », ce qui explique la possibilité d'employer village dans *village global*, à la

[41] B. Fradin (tirant partie des notions développées par Putnam 1975 sur la signification linguistique a proposé de ramener le sens des mots à des phrases génériques stéréotypiques. On pourrait ainsi associer à *village* des phrases comme « un village est un lieu convivial », « tout le monde se connaît dans un village ».



façon de Mac Luhan, par exemple. Il n’y a pas alors à invoquer une perte ou une érosion sémantique, car ce sont les mêmes valeurs qui permettent au mot de s’appliquer à toutes sortes de domaines d’expérience et d’être mobilisé pour un village campagnard, un village urbain, voire mondial. *Village* exprime donc fondamentalement l’idée d’une petite communauté liée à l’habitat dont la stabilité et les dimensions – plus restreintes que celles des grandes métropoles – permettent à des relations de proximité, de convivialité et d’entraide de s’établir entre ses membres.

Ce premier balayage reste rapide et en fonction de l’élargissement du corpus, il faudra sûrement revenir sur le caractère partagé des représentations de la ville. En fait, seuls certains quartiers de Paris, essentiellement le 7e, accessoirement le 14e, ont suscité cet emploi. Il sera intéressant de vérifier si ce motif est réservé à certains quartiers et à certaines couches sociales.

6. Bibliographie

Académie française, 1835. *Dictionnaire de l'Académie française* (6e éd.). Paris : Imp. de l'Institut F. Didot, 2 vols.

Académie française, 1992-2005, *Dictionnaire de l'Académie française* (9e éd.). Paris : Imp. de l'Institut F. Didot, 2 vols.

Adam, Jean-Michel, 1992. *Les textes, types et prototypes : récit, description, argumentation, explication et dialogue*. Paris : Nathan.

Anscombre, Jean-Claude, 1995. « La nature des topoï », in J.-C. Anscombre (éd), *Théorie des topoï*. Paris : Kimé, p. 49-84.

Anscombre, Jean-Claude, 2001. « Le rôle du lexique dans la théorie des stéréotypes ». *Langages*, n° 142, p. 57-76.

Bakhtine, Mikhaïl M., (1997 [1924-1970]). *Esthétique et théorie du roman*. Paris : Gallimard.

Bakhtine, Mikhaïl M., (1984 [1920-1974]). *Esthétique de la création verbale*. Paris : Gallimard.

Bakhtine, Mikhaïl M./Voloshinov, Valentin N., (1977 [1929]). *Le marxisme et la philosophie du langage*. Paris : Minuit.

Barbéris, Jeanne-Marie, (éd.), 1994. *La ville : Arts de faire, manières de dire*. Montpellier : Langue et Praxis, Univ. P. Valéry.

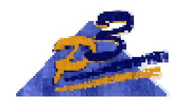
Barbéris, Jeanne-Marie, 2005. « La ville et ses composantes : l’émergence des catégories en interaction orale », in T. Bulot, L. Dubois (eds), « Signalétiques et signalisations langagières des espaces de villes », actes du colloque de Moncton (septembre 2005), *Revue de l’université de Moncton*, 31-60.

Bilger, Mireille, éd., 2000. *Corpus, méthodologie et applications linguistiques*. Paris : Champion.

Blanche-Benveniste, Claire, 1997. *Approches de la langue parlée en français*. Paris : Ophrys.

Blanche-Benveniste, Claire, 1995. « Quelques faits de syntaxe », in G. Antoine et R. Martin (éds). *Histoire de la langue française 1914-1945*. Paris : CNRS-Editions, p. 125-147.

Blanchet, Régis, 1997. *Un peuple-mémoire Les Roms*. Rouvray : Editions du Prieuré.



Boileau, Nicolas, 1666. *Satires du sieur D**** **[Texte imprimé] / (Despréaux).** Paris : C. Barbin.

Branca-Rosoff, Sonia, 2001. « La sémantique lexicale du mot “quartier” à l’épreuve du corpus *Frantext* (XII[e]-XX[e] siècles) ». *Langage et société* n° 91 “Espaces urbains, analyses lexicales et discursives”. p. 45-70.

Branca-Rosoff, Sonia, 2007. « Les usages discursifs du mot quartier : XVII[e]- XX[e] siècles », in P. Bacot et S. Rémi-Giraud (éds), *Mots de l’espace et conflictualité sociale.* Paris : L’Harmattan [coll. Langue & Parole], p. 169-185.

Bulot, Thierry, et Veschambre, V., 2004. « Sociolinguistique urbaine géographie sociale : hétérogénéité des langues et des espaces. » *Colloque international Espaces et société aujourd’hui (La géographie sociale dans les sciences sociales et dans l’action)*. Rennes.

Cadiot, Pierre, et Nemo, François, 1997. « Pour une sémiogenèse du nom ». *Langue française* n° 113, p. 24-34.

Calvet, Louis-Jean, 1994. *Les voix de la ville : introduction à la sociolinguistique urbaine*. Paris : Payot.

Cappeau, Paul, et Gadet, Françoise. 2007. « L’exploitation sociolinguistique des grands corpus. Maître mot et pierre philosophale ». *Revue Française de Linguistique Appliquée* 12/1, p. 99-110.

Chanet, Catherine, 2001. « 1700 occurrences de la particule *quoi* en français contemporain : approche de la « distribution » et des fonctions en discours ». *Marges linguistiques* n° 2, p. 56-80.

Charaudeau, Patrick, et Maingueneau, Dominique (éds), 2002. *Dictionnaire d’Analyse du Discours.* Paris : Seuil (en particulier, article « Stéréotype »).

Collet, Anaïs, 2005. « Sociologie des acteurs de la gentrification des quartiers anciens centraux d'hier et d'aujourd'hui », in L. Bonnet et L. Bertrand, *Mobilités, habitat et identités* (Paris : INED [coll. Documents du travail]).

Collet, Anaïs, 2008. « Les “gentrifieurs” du Bas-Montreuil : vie résidentielle et vie professionnelle », *Espaces et Sociétés* n[os] 132-133 “La gentrification urbaine”, mars, p. 125-142.

Corblin, Francis, 1987. « *Ceci* et *cela* comme formes à contenu indistinct », *Langue française* n° 75, p. 75-93.

Danjou-Flaux, Nelly, 1980. « *Au contraire, par contre, en revanche* : une évaluation de la synonymie ». *Bulletin du centre d'analyse du discours* n° 4, p. 123-146.

DELIC, « Corpus du français parlé Conventions de transcription. » <http://sites.univ-provence.fr/delic/corpus/conventions.html>

Ducrot, Oswald, 1995. « Topoï et formes topiques », in J. C. Anscombre (éd.), *Théorie des topoï* ; Paris : Kimé, p. 85-100.

Ebel, Marianne, et Fiala, Pierre, 1983. *Sous le consensus, la xénophobie*. Lausanne, Institut de science politique (Mémoires et documents, 16).

Faye, Jean-Pierre, 1972. *Introduction aux langages totalitaires, Théorie et transformations du récit*. Paris : Hermann.

Fernandez, Jocelyne, 1994. *Les particules énonciatives dans la construction du discours*. Paris : PUF.

Flaubert, Gustave, 1923. *Œuvres complètes illustrées de Gustave Flaubert. Éd. du centenaire*. Paris : F. Sant'Andrea, L. Marcerou.



Foucault, Michel, 1966. *Les Mots et les choses. Une archéologie des sciences humaines*. Paris : Gallimard.

Fouché, Pierre, 1969. *Traité de prononciation française*. Paris : Klincksieck.

Fuchs, Catherine et Le Goffic, Pierre, 2005. « La polysémie de *comme* » in O. Soutet (éd.), *La Polysémie*. Paris : Presses universitaires de Paris-Sorbonne. <http://halshs.archives-ouvertes.fr/halshs-00067939/en/>

Gadet, Françoise, 1997. *Le français populaire*. Paris : PUF (coll. Que sais-je).

Gadet, Françoise, 2006. « Hier comme aujourd'hui. Quelques faits de variation en syntaxe » in G. Williams, p. 31-45.

Galatanu, Olga, 1999. « Le phénomène sémantico-discursif de déconstruction reconstruction des topoï dans une sémantique argumentative intégrée ». *Langue Française* n° 123, p. 41-51.

Groupe d'Analyse des Données Textuelles (JADT.org). « Glossaire pour la statistique textuelle. » <http://tal.univ-paris3.fr/wakka/wakka.php?wiki=glossairec>

Habert, Benoit, 2005. *Instruments et ressources électroniques pour le français*. Gap/Paris : Ophrys (coll. L'essentiel français).

Habert, Benoit, Nazarenko, A., et Salem, A., 1997. *Les linguistiques de corpus*. U Linguistique. Paris : Armand Colin/Masson.

INSEE, « Statistiques locales. » <http://www.statistiques-locales.insee.fr/esl/accueil.asp>

Jaulin, Béatrice, 2000. *Les Roms de Montreuil*. Paris : Autrement.

Krieg-Planque, Alice, 2009. *La notion de « formule » en analyse du discours. Cadre théorique et méthodologique*. Besançon : Presses universitaires de Franche-Comté.

Labov, William, 1976. *Sociolinguistique*. Paris : Minuit.

Lakoff, Georges et Johnson, Mark, 1985. *Les métaphores de la vie quotidienne*. Paris : Minuit.

Lefeuvre, Florence, 2005 : « Le pronom *quoi* renvoie-t-il à de l'inanimé ? ». *Le Français moderne* n° 73/2, p. 170-183.

Lefeuvre, Florence, 2006 : *Quoi de neuf sur quoi ? Etude morphosyntaxique du mot* quoi. Rennes : PUR.

Legallois Dominique, 2003. « Essai sur la temporalité et le rythme du signe linguistique ». *Langage* n° 150, p. 48-60.

Le Goffic, Pierre, 1993. *Grammaire de la phrase française*. Paris : Hachette

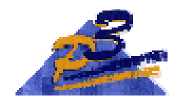
Lévy-Vroelant, Claire, et Segal, Jérôme, 2003. « Les Roms de Montreuil et d'ailleurs. Des immigrés européens particuliers. » *Les Temps Modernes* 624, p. 223-230.

Littré, Emile, 1863-1872. *Dictionnaire de la langue française*. Paris : Hachette.

Lodge, R. Anthony, 1993. *French: from dialect to standard*. London and New York: Routledge.

Lodge, R. Anthony, 2004. *A sociolinguistic history of Parisian French*. Cambridge : Cambridge University Press.

Martinez, William, et Zimina, Maria, 2002. « Utilisation de la méthode des cooccurrences pour l'alignement des mots de textes bilingues », in *Actes JADT'2002*, Journées Internationales d'Analyse Statistiques des Données Textuelles, St Malo. <http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2002/tocJADT2002.htm>



Martinez, William, 2003. *Contribution à une méthodologie de l'analyse des cooccurrences lexicales multiples dans les corpus textuels*. Thèse de Doctorat en Sciences du Langage, Université de la Sorbonne nouvelle - Paris 3, sous la direction d'André Salem. <http://williammartinez.fr/coocs/page.php?P=1&L=1>

Martinez William, et Leblanc, Jean-Marc, 2006. « *L'analyse contrastive des réseaux de cooccurrence Le monde dans les discours des présidents de la Cinquième République* », in Actes JADT'2006, Journées Internationales d'Analyse Statistiques des Données Textuelles, Besançon. < www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2006/ tocJADT2006.htm>

McEnery, Tony, Xia, Richard et Tono, Yukio. 2006. *Corpus-based language studies. An advanced resource book*. New York: Routledge Applied Linguistics.

Milroy, Lesley et Milroy, James, 1992. « Social networks and social class: Toward an integrated sociolinguistic model. » *Language in Society* n° 21, p.1-26.

Moirand, Sophie, 2004. « L'impossible clôture des corpus médiatiques ou la construction des observables entre catégorisation et contextualisation. » *Tranel* n° 40, p. 71-92, Université de Neuchâtel.

Morel, Mary-Annick et Danon-Boileau, Laurent, 1998. *Grammaire de l'intonation*. Paris : Ophrys

Noailly, Michèle, 1990. *Le substantif épithète*. Paris : PUF (coll. Linguistique Nouvelle).

Péchoin D. et Dauphin B., 1998. *Dictionnaire des difficultés du français d'aujourd'hui*. Paris : Larousse.

Pinçon-Charlot, Monique, et Pinçon, Michel, 2004. *Sociologie de Paris*. Paris : La Découverte (coll. Repères).

Préteceille, Edmond, 1995. « Division sociale de l'espace et globalisation. Le cas de la métropole parisienne. » *Sociétés Contemporaines* n° 22-23, p. 33-67.

Rastier, François, 1987. *Sémantique interprétative*. Paris : PUF.

Rastier, François, 2005. « Enjeux épistémologiques de la linguistique de corpus », in Williams (ed.) p. 31-46.

Recherches Sur le Français Parlé (RSFP) 2004. « Autour du corpus de référence du français parlé. »

Rossari, Corinne, 1997. (1ère éd. 1994). *Les Opérations de reformulation.* Berne : Lang

Schepper, Ewald, 1987. « Zur Mikrosystematik der transkategorialen plurifunctionalität von QUOI », in Dietrich, Gauger, Geckeler Hrsg., *Grammatik und Wortbildung romanischer Sprachen* (Tübingen : O. Narrverlag).

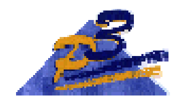
Siblot, Paul, 1997. « Nomination et production de sens : le praxème ». *Langages* 127, p. 39-55.

Siblot, Paul, 2007. « Nomination et point de vue : la composante déictique des catégorisations lexicales », in G. Cislaru, O.Guerin, K. Morim, E. Nee, T. Pagnier, et M. Veniard (éds) *L'acte de nommer – Une dynamique entre langue et discours* (Paris : Presses de la Sorbonne Nouvelle), p. 25-38.

Sinclair, John, 1991, *Corpus Concordance Collocation*. Oxford : Oxford University Press.

Thibault, Pierrette et Vincent, Diane, 1990. *Un corpus de français parlé*. Québec : CIRAL.

Trésor de la langue française. <http://atilf.atilf.fr/tlf.htm>



Vincent, Diane, 1993. *Les ponctuants de la langue et autres mots du discours*. Québec : Nuit Blanche.

Vincent, Diane et Dubois, Sylvie, 1997. *Le Discours rapporté au quotidien*. Québec : Nuit Blanche.

Williams, Geoffrey (éd.), 2006. *La linguistique de corpus*. Rennes : Presses universitaires de Rennes.

Yaguello, Marina. 1998. *Petits faits de langue*. Paris : Le Seuil.